Metaphor Through an Evolutionary Perspective on Meaning

Ehud Rahat

Ph.D.
University of Edinburgh
1990

Declaration

I declare that this thesis has been composed by myself and that the research reported therein has been conducted by myself unless otherwise indicated.

Edinburgh, 15 July 1990.

Ehud Rahat

Acknowledgments

First and foremost I wish to thank my dedicated supervisor, Dr. David Bloor. David has spent endless time with me in discussions of my work, more than anyone can expect from a supervisor. He has helped me with many useful ideas, encouraged me in difficult moments, and managed to guide my work without a trace of imposing his own inclinations on anything that I had brought up in discussions. I am very grateful to him for his competent instruction. I am also grateful to Dr. Paul Schweizer, my second supervisor, who made helpful comments on my writings.

I thank Jerry Seligman for many stimulating discussions, that helped me clarify my sometimes confused ideas. I also wish to thank various philosophers that very kindly have responded to my questions and commented on my work: special thanks to Prof. Millikan and Prof. Cooper, but also to Prof. Fodor, and Prof. Davidson. I am grateful to Richard Cooper who answered many nagging questions concerning the formatting of this thesis.

Many thanks to the various bodies that have financially supported me during my studies: The B'nei B'rith organization, the British Council, and the Universities and Colleges of the United Kingdom which granted me the ORS Award.

Finally, I wish to dedicate my work to my beloved parents, and to the joys of my life, Ayelet and Shye.

Abstract

This thesis applies a theory of meaning, suggested by Millikan (1984), to explain the problematic notion of metaphorical meaning. There is a persistent tension in existing accounts of metaphorical meaning: on the one hand, the fact that metaphors are novel and creative uses of words suggests that no systematic account can be given as to what metaphorical expressions mean. On the other hand, the fact that metaphors are used intelligibly in all areas of discourse, and that metaphorical interpretations can be distinguished from nonsensical interpretations, indicates that there is a systematic relation between the words used and what they are said to mean metaphorically. I exemplify this tension by discussing three accounts of metaphor offered by Searle, Davidson, and Hesse.

I then present Millikan's theory of meaning that is based on an analogy she makes between linguistic and biological categories. Her basic claim is that linguistic entities can be ascribed with a proper function, just as biological entities are so ascribed, according to their evolutionary history. The proper function of a device is that function which earlier tokens of that device had performed, which (partly) accounts for the proliferation of that device. The proper function of a linguistic device, and the conditions that normally obtain *when the device performs its function*, determine the meaning of the device.

Next, I counter Fodor's criticism of Millikan's approach. I show that the conditions which Millikan requires for attributing a meaning to a symbol satisfy, in many cases, Fodor's own suggestion for such conditions. His criticism must therefore apply to his own theory. I show, however, that it is Fodor's theory, rather than Millikan's, that suffers from a related problem to that which he has challenged her with.

Finally, I apply Millikan's theory to the problem of metaphorical meaning.

First, I discuss Millikan's own conception of metaphor, which is apparently similar to Searle's. I argue that her approach nevertheless manages to withstand most of the criticisms that apply to Searle's theory. Next, I develop an alternative account of metaphor, based on Millikan's theory of meaning. My claim is that in the analogy between linguistic and biological categories, metaphor is analogous to cases of mutation in biological systems: they are the mechanism by which linguistic changes are introduced. To answer the question of metaphorical meaning, then, we need to ascertain the proper function of mutants. This proper function, I claim, is the introduction of adaptive changes to existing devices. Analogously, metaphors have the proper function of introducing semantic changes to the words they are comprised of.

Two results follow from the analogy: first, an important insight gained from Millikan's theory is that any generalizations made about the structure and functioning of a device need only apply to those devices that do perform their proper function. Accordingly, any claims about systematicity in metaphorical meaning need only apply to metaphors that perform their proper function of introducing semantic change. That is, only to those metaphors that have become established in language. Conversely, not every 'one-off' metaphor needs to be accounted for. Second, evolution theory cannot be used to pre-determine what particular function a device may acquire through selection, but nevertheless it can characterize constraints on the acquisition of such functions. These constraints are determined by the evolutionary history of existing devices. Similarly, I claim that no account can be given that will uniquely determine the meaning that a metaphorical expression may acquire in the process of its becoming an established part of a language. This reflects the first aspect of the tension described above, concerning the novelty and indeterminacy of metaphorical meaning. On the other hand, I argue that *constraints* on meaning change can be identified. This acknowledges the intuitions that metaphorical meaning is structured. The notion of metaphorical meaning should be understood as applying to that range of possible mean-

ings, determined by the constraints referred to above, which a metaphor may acquire as it becomes established. This understanding of the notion coherently captures the two apparently opposing characteristics of metaphor.

Contents

Introduction **1**

I Metaphor and Systematicity **5**

1 Searle on Metaphor **6**

1.1 Exposition 6

1.2 Discussion 12

1.2.1 Speaker's utterance meaning 14

1.2.2 Comprehension vs. meaning 20

1.2.3 Searle vs. Grice 24

1.2.4 Are metaphors systematic? 28

1.2.5 Conclusion 32

2 Davidson on Metaphor **33**

2.1 Metaphors and Malapropisms: Davidson on the Limits of the Literal 33

2.1.1 Introduction 33

2.1.2 Literal meaning and communication 36

2.1.3 Metaphors and malapropisms 41

2.2 Discussion 52

2.2.1 Davidson's view on systematicity in metaphor 53

2.2.2 The basic dilemma again 54

2.2.3 Settling the dilemma 55

2.2.4 Re-evaluating Davidson's argument 57

2.2.5 Meaning, or just content? 60

3 Hesse on Metaphor **65**

3.1 Exposition 66

3.2 Discussion 70

3.2.1 The metaphoric version of the deductive model of scientific explanation. 72

3.2.2 An alternative interpretation of Hesse's view on metaphor. 77

3.2.3 Hesse vs. Davidson. 81

3.3 Conclusion 83

II An Evolutionary Look at Meaning **85**

4 Millikan's Theory of Meaning **86**

4.1 Introduction 86

4.2 Theory of Proper Functions 91

4.2.1 Direct proper functions 91

4.2.2 Derived proper functions 96

4.3 Three aspects of meaning 100

4.3.1 Stabilizing proper function as the first aspect of meaning 101

4.3.2 Sense as the second aspect of meaning 110

4.3.3 Intension as the third aspect of meaning 116

5 A Reply to Fodor: Who Really Solves the Disjunction Problem? 120

5.1 Introduction 120

5.2 Two naturalistic theories of meaning 121

5.3 The Disjunction Problem 127

5.4 Some problems with Fodor's arguments 129

5.5 How does Fodor differ from Millikan? 133

5.6 Who solves the disjunction problem? 141

6 Millikan's Account of Metaphor 144

6.1 Introduction 144

6.2 Metaphorical meaning as derived proper function 146

6.3 Examining Millikan's account of metaphor 151

6.3.1 Conventional metaphors 152

6.3.2 Fresh metaphors 154

6.3.3 Cooper's objections 156

7 An Evolutionary Perspective on Metaphor **160**

7.1 Introduction 160

7.2 Biological mechanisms of variance 162

7.2.1 The evolutionary history of variants 162

7.2.2 Mechanisms of stability 165

7.2.3 Mechanisms of variance 167

7.2.4 The proper function of variants 168

7.3 Metaphor as a semantic mechanism of variance 172

7.3.1 The proper function of fresh metaphors 172

7.3.2 Separating functioning from non-functioning metaphors . 176

7.3.3 Identifying the Normal characteristics of functional metaphors 179

7.3.4 Metaphorical meaning 182

7.3.5 Metaphorical clusters 186

7.4 Conclusion 187

References **190**

Introduction

The central question that this thesis is set to answer is the following: is there a systematic relation between a linguistic expression's literal meaning, and something that may be reasonably called it metaphorical meaning? This question is close to the question 'what do metaphors mean?', but I prefer to consider the above formulation. What meaning is, is a very controversial issue in philosophy, and asking what metaphors mean invites too many assumptions to be made already in the way the question is understood, let alone in answering it. Considering these assumptions cannot be avoided, but my formulation is intended as a buffer to these assumptions: a question about systematic relations in general can perhaps be answered in terms that will not depend, at least to begin with, on a particular notion of meaning. Put another way, I think that the question about systematicity can be dealt with in a way that sets aside particular differences between theories of meaning. Moreover, I feel that this issue of systematicity is the important question that underlies the discussion about what metaphors mean, and that aspects of meaning other than that are a matter of less substance.

The question asked poses a dilemma in trying to answer it. Intuitively, metaphors are somehow more structured than non-sense expressions, but not as structured as literal expressions. More in particular, if one accepts that there is a systematic relation between an expression and its metaphorical meaning, one

loses the main reason for distinguishing metaphorical meaning from literal meaning. Also, metaphor implies innovation and creativity, a breaking away from following given linguistic rules. If one thinks that the metaphorical meaning of an expression can be captured by systematic rules, one is effectively denying metaphor the character of innovation. Finally, there are just too many examples of metaphors the interpretation of which seems to elude any kind of rules. On the other hand, if one denies systematicity, this raises other worries: can *any* interpretation of an expression be legitimately considered a metaphorical interpretation? What is the difference between understanding a metaphor and *mis*understanding it? And how are metaphors understood, if there is nothing to guide their interpretation? Finally, again, many metaphors do seem to display systematicity in their interpretations, and it seems inappropriate to dismiss this.

David Cooper, in his book 'Metaphor' (1986), discusses three approaches to metaphorical meaning. The first he labels 'the Traditional view', which argues that metaphors do have a meaning that is determined by the expression used metaphorically. The second is labelled 'the Standard view', and is presented through Searle's account of metaphor. Searle argues that metaphors do not have, in themselves, a metaphorical meaning, but that metaphorical meaning is what the speaker means when uttering a metaphor. The third view claims that metaphors do not have a metaphorical meaning, and this is the view Cooper argues for, following Davidson.

In the first part of the thesis I will follow Cooper's tripartite division (though not in this order), to demonstrate the dilemma in answering the opening question. I start with Searle, and argue that his account fails because it tries to have it both ways, i.e. to let metaphors be systematic and non-systematic at the same time. The second chapter discusses Davidson's account, which opts for the 'non-systematic metaphor' option, together with other reasons for why metaphors cannot have a metaphorical meaning. I try to answer those other

reasons, and argue against his choice of non-systematicity. Finally, I present Hesse's account of metaphor. She argues for what seems to be a systematic account of metaphorical meaning. I show that this account is unacceptable (I add, however, a re-interpretation of her original account, based on a later paper that she's written, to show that it doesn't really have the flaws that I've pointed out, because it is not exactly a choice of the 'systematic metaphor' possibility). Throughout the chapters in the first part I begin to outline the general character of a solution to this basic dilemma. I argue that metaphorical meaning is *partially* systematic in its relation to the literal meaning of an expression. That is, that an expression's literal meaning sets constraints, without uniquely determining its metaphorical meaning.

Stated in this way, of course, this solution is both formalistic and unrevealing: it lacks substance. To provide that substance we need a model of what happens when metaphors are used; we should then show how this model satisfies the general conditions that we have placed on a solution to our problem. This is the role played by the second part of the thesis. In the second part, I present a theory of meaning that was suggested by R.G. Millikan (1984), and based on an analogy between language devices (e.g. words and sentences) and biological devices (e.g. eyes, hearts, kidneys, etc.). The proper function of those devices is identified with the aid of evolution theory, and the meaning of language devices is explicated in terms of their function. I go on to defend this theory from its main critic, Jerry Fodor.

In the last two chapters, I tie together the two themes of the thesis—systematicity in metaphorical meaning, and the evolutionary account of meaning. I aim to show how applying Millikan's account to the original question about metaphor answers the basic dilemma. It answers it, first, by offering an explanation of metaphor that agrees with the solution I have outlined, of partial systematicity in metaphorical meaning. Second, and most importantly, it shows that claims

about systematicity in metaphorical meaning need not apply to *all* metaphors. Instead, it provides a principled way of distinguishing a sub-class of metaphors, to which the claims about the (partial) systematicity of metaphorical meaning should properly apply to. This explains the confusing presence of examples both of systematic, and of non-systematic metaphors.

Part I

Metaphor and Systematicity

Chapter 1

Searle on Metaphor

This chapter will begin with a short exposition of Searle's view on metaphor. In what follows, I will argue that his account only restates the problems that metaphors raise, without offering any real solutions. This is exactly why I have chosen to present his work first: to have the problems (re)stated. So, let us see what Searle has to say on metaphor[1].

1.1 Exposition

The question about metaphor that Searle sees as most important is 'how metaphors work?'. This is a special case of explaining how communication succeeds when what the words or sentence that a speaker uses mean one thing, yet the speaker intends to communicate something else. Other such cases are indirect speech acts and irony (cf. 'Indirect Speech Acts', in Searle 1979). Presenting the problem in this way is based on a distinction that Searle makes between *word*, or *sentence meaning*, and *speaker's utterance meaning*. The first refers to the meaning that linguistic expressions have as types of expression; it is those truth

[1]Searle's view on metaphor is presented in his book 'Expression and Meaning' (Searle 1979). All page references in this chapter, unless otherwise stated, are to this paper.

conditions that are determined by elements that are realized in the semantic structure of a sentence. The latter refers to what a speaker intends to convey by uttering a particular expression. What we intuitively call the metaphorical meaning of an expression, that which is different from its literal meaning, is the speaker's meaning. Literal meaning is identical with sentence meaning, and literal use is one where the speaker intends to convey what the sentence means. The meaning of the expression used metaphorically (i.e. its sentence, or word meaning) is not affected by a single deviant use, where a speaker chose to use it to convey something other than what it means. This is not to say that an expression may not acquire, over time, a meaning that it didn't have before, and that coincides with the speaker's meaning on a particular occasion of its use. But when it does, using it to convey that new meaning will cease to be metaphorical. A metaphor is a metaphor only to the extent that speaker's meaning differs from the expression's own meaning at the time of utterance. According to Searle, when the metaphor 'dies', i.e. becomes established, it becomes an expression with an ambiguous literal meaning.

The fact that speakers are able to communicate to their hearers what they mean when they use metaphorical utterances (as well as ironical utterances and indirect speech acts) indicates, Searle claims, that there are some principles shared by the speakers and hearers which enable non-standard communication. The task of a theory of metaphor is to spell out those principles. Doing this will reveal in what way metaphors are different from other forms of non-literal communication, where other principles are used for recovering what a speaker meant by his utterance.

The claims made so far may seem to imply a position that takes sentence meaning to be context independent, in contrast with the intended metaphorical meaning that depends on the context at least in so far as it depends on the speaker. Searle explicitly rejects this view. What a sentence means does indeed

depend on the context of its use, even when it is used literally, and thus may vary with the context. That this is the case is seen already when considering sentences that contain indexicals, in that the truth conditions they determine are context dependent. For example, the utterance of the sentence 'the cat is on the mat' will define truth conditions that will depend on the time of utterance, because of the present tense, and on the contextually dependent definite descriptions 'the cat' and 'the mat'. How the indexicals are affected by the context can be specified in advance of their use on a particular occasion, and their presence is realized in the semantic structure of the sentence. But, there are many more contextual determinants of truth conditions that are not so realized, and that cannot in principle be specified in advance. Searle calls these additional factors 'background assumptions'. In our example, such a background assumption may be the gravitational field relative to which the claim that the cat is on the mat is made true. In the normal setting one has in mind this is clear, but what the truth conditions would be for a cat and a mat floating in outer space is not so clear. Searle is explicit in discussing his view on literal meaning, so it is made clear where *not* to look for differences between literal and metaphorical uses. They are not distinguished just by their sensitivity to the context of utterance[2]. If context sensitivity is to play a role in the account of metaphor, it has to be specified how this sensitivity differs from that of literal uses.

To make the discussion of metaphors and their related speaker's meaning clear, Searle confines it to simple sentences of the subject-predicate form. The observations on the simple cases can presumably be generalized to deal with the more complex. A case of simple metaphorical use of a sentence is one in which a speaker utters 'S is P', but intends to communicate that S is R, where P and R are different. The actual examples that Searle presents are of 'dead', or established metaphors, that by his own standards aren't really metaphors but

[2]For a full account of the role of contextually fixed background conditions see 'Literal Meaning', in (Searle 1979).

just expressions with relatively new or additional meanings. He claims that this poses no problem, as such metaphors were at one stage new, and that they are particularly interesting because 'their continual use is a clue that they satisfy some semantic need' (p. 83). Among his examples are the following: the metaphor

(1) (MET) Sally is a block of ice

corresponds to the paraphrase

(1) (PAR) Sally is an extremely unemotional and unresponsive person

and

(2) (MET) Richard is a gorilla

(2) (PAR) Richard is fierce, nasty, and prone to violence.

The use of paraphrases to express the metaphorical meaning creates an uneasiness, as it may be felt that something is lost if the metaphor is expressed by a literal sentence. Searle acknowledges this problem, and says that what is lost, and why, needs to be explained. Nevertheless, 'the paraphrase or something like it must express a large part of speaker's utterance meaning, because the truth conditions are the same [for the paraphrase and the speaker's meaning]' (p. 83). Sometimes we may not be able to find an exact paraphrase to a metaphor, even though we may understand it very well. An example for such a case would be

(3) (MET) 'The ship ploughed the sea'.

This may happen because there is no literal expression in the current language that conveys what the metaphor does, and this is exactly why a metaphor is needed: to fill such semantic gaps. In other cases still, a metaphor may have an indefinite range of possible meanings, as when Romeo says

(4) (MET) Juliet is the sun.

We may realize that this utterance is intended to express admiration of some sort, but we are unable to restrict the interpretation to one particular proposition.

The original question that Searle has set out to answer, 'how metaphors work?', can now be formulated more precisely as 'how is it possible for a speaker to say 'S is P' and mean 'S is R'?'. Or, from the hearer's point of view, how can he conclude from hearing 'S is P' that what is meant is 'S is R'?[3] Searle specifies the steps of reasoning that the hearer must go through, but says that these are not necessarily steps that he follows consciously. Instead, what he describes is 'a rational reconstruction of the inference patterns that underlie our ability to understand such metaphors' (p. 104). Searle stresses that the relations that hold between the uttered sentence, 'S is P', and the intended one 'S is R', are not meaning relations, i.e. the one doesn't *mean* whatever the second does. Presumably, the second is implied, in some sense, by the first. This idea is similar to Grice's familiar notion of implicatures—those things one may correctly deduce from the meaning of what had been said to him (cf. Grice 1975). These relations, Searle point out, are not part of the implied meaning, which is just 'S is R', nor are they part of the meaning of what has been said.

The first step a hearer takes is an attempt to understand the uttered sentence

[3]Searle seems to take these two formulations as equivalent. However, the latter implies successful communication, while the former doesn't necessarily. A speaker may mean something but not be understood as intended.

literally. This will usually reveal that the utterance is defective in some way, such as being false, nonsensical, or irrelevant to the ongoing discourse. There will be, in other words, some clue that a different meaning from that of the sentence should be sought for. This will lead to the second step, that will involve finding possible values for an R, to put instead of the P term, that will rectify the defectiveness of the sentence uttered. There are various principles for computing such values for R, that are all dependent on, or are a function of, the P term. The main task for a theory of metaphor is to specify those principles which I shall call in short PMIs (Principles of Metaphorical Interpretation). We will soon look at some of the principles that Searle suggests, but in general they are such that they offer a wide range of possible R values, not all of which can be applied to the subject of the sentence, S. This leads to the third step, in which the possible values of R are narrowed down to the relevant ones by looking at the S term and seeing which of these values are likely, or even possible properties of S.

One of the PMIs will be the following: 'look for salient, well known, and distinctive features of P things as possible values for R'. We can now demonstrate with an example how following the above steps will lead to recovering the metaphorical meaning of an utterance. Suppose someone utters 'Sam is a pig'. In the first step, the hearer will realize that the literal meaning of this sentence is defective, because the person talked about clearly couldn't be a pig. This will initiate the second step, of finding possible values that can replace 'pig'. The principle we have mentioned will lead the hearer, using his factual knowledge about pigs, to come up with features such as being fat, gluttonous, slovenly, filthy, but also having a distinctive shape, having a short, curly tail, and living in the mud. In the third step, the hearer will check which of these features may apply to Sam. He may then reject the latter group of features (i.e. the shape, tail, etc.) , and will remain with the interpretation that Sam is fat, filthy, and so on.

The first PMI involved features that are typical of P. Another PMI will invoke things that are P by definition, e.g. (MET) 'Sam is a giant' will be interpreted as (PAR) 'Sam is big' because giants are, by definition, big. These first two PMIs depend on features that Ps really have. But sometimes a metaphor may involve alluding to features that are not actually true of P, but instead are just commonly associated with it. Thus another principle says that we should look for features of P that are often said or believed of it even if not true. For example, (MET) 'Richard is a gorilla' will be paraphrased as 'Richard is mean, nasty, and prone to violence' even though the speaker and hearer may both know that in fact gorillas are shy, timid and sensitive creatures. Similarly, this principle will explain the understanding of metaphors that involve reference to objects that do not exist, e.g. 'Cathy is a witch'. This avoids problems that theories of metaphor which assume a comparison between two actual objects find difficult to explain.

There are several other principles that Searle presents, and he assumes that there are more to be revealed. However, he summarizes his position by stating that employment of the three step strategy and the PMIs are necessary and sufficient conditions for communication via metaphorical expressions. The combination of this strategy and the principles of interpretation will be labelled SPMI, for short.

1.2 Discussion

The main problem about metaphors, as I see it, is a tension that is created by trying to explain the connection between the (literal) meaning of the words used in a metaphorical utterance, and whatever we care to call the metaphorical meaning which is different from that literal meaning. On the one hand, this connection seems new, and specific to that occasion of utterance. On the other hand, there does seem to be something systematic and predictable about this

connection, that distinguishes a metaphor from nonsensical strings of words, or plain mistakes about how words should be used. We must keep in our mind that people who utter metaphorical expressions deliberately choose the words they do rather than being confused about what the words they choose mean. If we look at the definition of 'literal' in the OED we find that it means 'taking words in their usual or primary sense and applying the ordinary rules of grammar, without mysticism or allegory or metaphor'. Literal is defined as usual and ordinary, and is contrasted with metaphor, which implies its being unusual in some sense. Admittedly, this formulation allows for the possibility that in literal utterances people use an ordinary set of rules, whereas with metaphor they may be using some special rules, that is, rules for constructing and interpreting metaphors. But I think this possibility is not quite satisfactory, as intuitively metaphors seem to involve novelty that requires a certain amount of creativity over and above that of just employing a different kind of rules.

Those two opposing characteristics of metaphor are acknowledged by Searle in two aspects of his theory. Its particularity, its being unusual and irregular, is echoed in his claim that metaphorical meaning is speaker's utterance meaning, and his saying that it is a mistake to try and locate a metaphor's meaning in the words uttered. Word, or sentence meaning, presumably has to do with there being an underlying regularity in the way expressions are interpreted, whereas the assumed creativity that metaphorical utterances involve should block any appeal to regularities in accounting for their interpretation. But Searle says on the other hand, that 'The relation between the sentence meaning and the metaphorical meaning is systematic rather than random or ad-hoc', a conclusion that must follow, he thinks, from the fact that speakers can communicate using metaphorical utterances.

In the following I will show that Searle, in his account, doesn't notice that there is a contrast between these two aspects, and doesn't resolve the tension

between them. He seems to just pay lip service to the idea that there is something irregular about metaphors by calling their meaning 'speaker's utterance meaning'. Most of his account then ignores this aspect and shows how metaphors are in fact systematic and regular. In that, he makes the distinction he cherishes, between speaker's meaning and sentence meaning, one that is not really very interesting: if one can reliably reveal the meaning of a metaphor without really considering the speaker's intentions or anything else particular to the specific occasion of utterance (except for general background assumptions), what is the point in separating the metaphorical meaning from the words and assigning them to the speaker? My argument will be of the following form: first, I will show that the idea that metaphors mean whatever their speaker intends them to mean is not consistent with the claim that metaphorical meaning has a systematic relation to the word, or sentence meaning. Second, I will argue that Searle's claim that this systematicity in metaphorical meaning applies to the comprehension processes involved in the understanding of metaphors rather than to their meaning is both wrong and unnecessary. Third, I will distinguish Searle's view on metaphor from Grice's treatment of conversational implicatures, in order to clarify the fact that though the treatments are similar, my criticism of Searle doesn't apply to Grice's account. Finally, I will discuss doubts as to whether any systematicity of the kind Searle has claimed to reveal does exist between the literal and the metaphorical meaning of metaphors.

1.2.1 Speaker's utterance meaning

David Cooper devotes a good part of his book 'Metaphor' (Cooper 1986) to criticize Searle's account. The following is one of his arguments against it, which he calls 'the indeterminacy objection'. I present it here both for its intrinsic force as an argument against Searle's thesis, but more importantly for a general point it makes about the understanding of the notion of speaker's utterance

meaning. Cooper focuses our attention on a phenomenon that Searle discusses in passing: that of a metaphor with an indeterminate interpretation, i.e. 'one which admits of more than one interpretation, none of which can be demonstrated as uniquely correct' (*ibid*, p. 70). He argues that Searle's account 'is forced to give a mistaken explanation of indeterminacy' (*ibid*, p. 70). There are three ways to understand Searle's appeal to speaker's meaning as explaining indeterminacy. An indeterminate metaphor is either (i) a case where we are not sure what the actual speaker meant by his utterance, (ii) a case where the speaker actually meant something indeterminate, or (iii) a case where possible speakers may have meant different things by uttering that metaphor.

Considering the first possibility, it should be noticed that many times we interpret metaphors without even knowing who their speaker is, or with knowing very little about him. That means we are in fact speculating about what a possible speaker may have meant by uttering those words, which is a case of the third kind mentioned above. Some cases, then, must be explained by that third option. But even when we do know exactly what the speaker meant (suppose he told us), we are free to interpret his words in several ways that might make the interpretation indeterminate. Indeterminacy of metaphorical interpretation, we see, is not affected by knowing the speaker's intentions either way.

To the second possibility, that the speaker meant something indeterminate, Cooper replies that the point made against the first option applies here as well: speaker's intentions, determinate or not, do not affect indeterminacy. Also, it is pointed out what indeterminacy involves in this case: it is not that a speaker means several things at once, or one thing at one time, another on a different time. These would be cases of determinate, though multiple, metaphorical meanings. Instead, the speaker must mean something that is itself open-ended. This, Cooper claims, is an exceptional case, that surely doesn't cover all cases where we wish to say that a metaphor is indeterminate.

Finally, we have the option that Searle is talking of what possible speakers may mean. Understood one way, this trivializes the explanation and makes it a truism: possible speakers may mean anything by uttering a sentence, and thus every metaphor is indeterminate in this sense. Even more problematic is that such an account allows metaphorical meanings to be associated with an utterance in a way that has nothing to do with the words uttered, which is absurd. An adequate account of metaphor must, in some way, establish a connection between the metaphorical meaning and the meaning of the words used. Alternatively, we may understand this third option as saying that indeterminacy is explained by what possible speakers may *reasonably* mean by uttering an expression. But then, we are assuming some standard of reasonableness relative to which possible speaker's meaning is judged. The appeal to speaker's meaning is made redundant, as the expression's meaning is then determined by those standards, not by what any speaker actually means.

Cooper has made his valid objection to Searle focusing on the problem of indeterminate metaphors, but applying similar lines of argumentation reveal further general problems internal to Searle's account. The problems are due, basically, to the clash between the idea that a metaphor means what the speaker intends it to mean, and the specification of principles that should reveal what the speaker means. Speaker's intentions are inherently free of any limitations—a speaker may intend to convey just anything by using a linguistic expression. Principles that try to explain, in a way that is not ad-hoc, how a speaker manages to achieve his intentions, can only apply to the means that he may choose for doing so; they cannot constrain what he may actually intend. Yet Searle claims that the principles he offers do reveal the metaphorical meaning of utterances. This, I will show, must have one of two consequences: either the principles are wrong, or he is not taking seriously the idea that metaphorical meaning is speaker's meaning.

What are the principles specifying, according to Searle? Let us look at his exact formulation:

> 'the following strategies and principles are individually necessary and collectively sufficient to enable speaker and hearer to form and comprehend utterances of the form 'S is P', where the speaker means metaphorically that S is R (where P is different from R)'.
>
> (p. 112)

Let us brake down this claim into its elements. First, Searle claims that the SPMIs are necessary and sufficient conditions *for a speaker to form* utterances of the above kind, where he means something different from what he says. This claim is obviously false: speakers do form utterances that do not conform with these principles. This is clear from the fact that they are not always understood, and this is not always due to failures of the hearers to follow these principles in interpretation. Many times, the speaker just doesn't choose the right words to convey his thought. This, among other things, is what makes a bad metaphor. Searle couldn't, then, claim that his principles invariably reveal speaker's intentions in communication.

Perhaps what Searle means is that the SPMIs are necessary and sufficient conditions for the conjunct, of speakers forming and hearers understanding such utterances where the speaker says something different from what he means to convey. In other words, what he offers must hold in all cases where communication is successful. To evaluate this claim, let us find out why Searle thinks that these principles, or in fact any other principles, should exist.

> 'In order that the speaker can communicate using metaphorical utterances, ironical utterances, and indirect speech acts, there must be some principles according to which he is able to mean more than, or something different from, what he says—principles known to the hearer, who, using this knowledge, can understand what the speaker

means'.

(pp. 77–8)

Searle is saying two things: one, that from the fact that communication succeeds it follows that there must be principles that govern the connection between what the speaker says and what he means. Two, that these principles must be shared by the speaker and the hearer. I would like to argue, following Davidson, that the first claim is probably false, or at least dubious, and that the latter is just wrong. Speakers do form utterances that do not mean what they intend to convey, and are understood, in ways that do not conform to these principles. Examples of such cases are slips of tongue, uses of unfamiliar idiolects, and malapropisms[4] (malapropisms are cases where a speaker uses one word in place of another, typically when the word used resembles the one that means what was intended). In such cases it is not at all clear that there are any general principles that characterize the interpretation process, as these divergent uses are randomly created. But even if we suppose that there are any such principles of interpretation in those cases, the principles are surely not shared by the speaker and hearer: the speaker has one reason for mispronunciation, say his speaking impolitely with food in his mouth (what principle guides him in such a case?), and the hearer understands him for a different reason, say by knowing from the context of utterance what the speaker was likely to want to say that sounds similar enough to what he has mumbled.

We have established that success of communication is not the reason for why any shared principles must apply to all cases of indirect communication, and if there are any principles, they must look *very* different from those Searle had offered. But suppose Searle draws our attention to the fact that, in what was quoted above, he only said that his principles must hold in cases 'where the speaker means *metaphorically* that S is R...'. Thus, while it may be true

[4]This point is derived from 'A Nice Derangement of Epitaphs' (Davidson 1986) and will be discussed in Chapter 2.

that communication does succeed in different cases for many different reasons, some of which may not involve any principles, this is not the case when metaphors are concerned. In fact, this is why we give this special form of indirect communication a special name, 'metaphor', to distinguish it from other forms such as irony, indirect speech acts, and maybe many others that cannot be classified because they are so messy. Such a claim can be understood in two ways: either as saying that metaphors are structured, as a matter of fact, in a way that is captured by the SPMIs, or that Searle defines metaphor as cases that can be explained by these principles.

Let us expand the first claim. It could be that Searle is not concerned with saying what metaphor is exactly. Suppose there is some adequate definition of which cases, when a speaker says 'S is P' and means 'S is R', are metaphors, but that this definition is yet unknown. All he is saying is that whatever it is, the SPMIs explain this odd phenomenon where people manage to understand, *when they do understand*, things that weren't said. This allows for the possibility that there are cases where someone utters what we may want to say is a metaphor, yet he is not understood because the SPMIs do not reveal his intended meaning. Can such an understanding of Searle's account let the SPMIs and the claim that metaphors mean whatever their speaker means live together? I think not. This claim depends on the assumption that there is a standard by which people's intentions are considered metaphorical or not. But this clashes head on with the idea that metaphors mean what their utterers mean, since as I explained above, people's intentions are inherently unconstrained. Also, the notion of 'speaker's meaning' is made redundant in the account. The metaphor means what it does because of the standards that say how P and R should be related. The SPMIs do reveal the speaker's meaning, but this is only because they approximate the definition that made his intended meaning a case of metaphor. This would be like giving principles for revealing a speaker's intentions when it is given that he spoke literally. Such principles would add nothing to the regular rules that

specify the meaning of the linguistic forms he uses.

The other possibility, that Searle has defined for us what a metaphor is, suffers from exactly the same problems as the first. Again, his principles would constrain the speaker's meaning, which cannot be constrained. And again, his principles would work not because they revealed the speaker's meaning, but because the speaker's meaning would match the definition.

To summarize the above argument: first, from the fact that people form sentences that do not express what they intend to communicate and are nevertheless comprehended it does not follow that there are any fixed principles that invariably govern the relation between what people mean and what they say. The understanding of metaphors, in particular, is not obviously guided by any rules. Second, if there are any principles that govern the particular case of metaphorical communication, then metaphors cannot be claimed to mean what their utterer intends them to mean, because they must mean what those principles determine that they mean.

1.2.2 Comprehension vs. meaning

I have claimed that Searle's specification of the SPMIs is not compatible with the unrestrictedness that is implied by relating metaphorical meaning to the speaker. We have seen that Searle himself, while not acknowledging this conflict, seems to give more significance to the existence of the principles than to the notion of speaker's meaning. His theory, therefore, seems quite similar to those he criticizes which 'try to locate the metaphorical element of a metaphorical utterance in the sentence or expressions uttered' (p. 77). It should be noted, however, that another difference that Searle points out between his theory and some of the others, stays intact even if we drop his claim about the metaphorical meaning being speaker's meaning. This is the distinction between the *meaning*

of a metaphorical utterance, or its truth conditions, and the *comprehension processes* and principles of inference that they involve. Discussing the role of similarity in accounts of metaphor, Searle criticizes other theorists for offering accounts in which 'the meaning of a *metaphorical* statement is always given by an explicit *statement* of similarity' (p. 98). Searle claims instead that 'though similarity plays a role in the *comprehension* of metaphors, the metaphorical assertion is not necessarily an *assertion* of similarity' (p. 88). For example, if (MET) 'Sam is a pig' means that (PAR) Sam is dirty, we see that neither the notion of similarity, nor the pig to whom Sam is presumably similar in some respect are part of the meaning, nor should they be part of it. Instead, similarity and the pig will have a role in explaining how one can arrive at the metaphorical meaning, i.e. they will be part of the explanation of the comprehension process.

I will not assess here the validity of the claim that the statement of similarity shouldn't be part of the meaning of metaphor. Also, I'm not concerned here with the question of the role of the relation of similarity in the account of metaphor. What does concern me is the important distinction that Searle points out between what a metaphor means, and the relation (be it similarity, or any other relation) this meaning has to the metaphorical sentence's literal meaning. I think that this distinction should be kept in mind, and any account of metaphor should make clear how it maintains the distinction.

I have a problem, however, with Searle's taking for granted that the relation between the two meanings, the literal and the metaphorical, must reside in the comprehension process. For reasons which follow from the first part of this discussion, I think that identifying these assumed relations as part of the comprehension process is both wrong, and an unnecessary complication. But before specifying my objections, I would like to comment on Searle's use of this notion of a comprehension process. Like his use of 'speaker's meaning', where it is not clear if he is talking of an actual speaker's meaning, a possible meaning of an

actual speaker, or an actual or possible meaning of a possible speaker, here too Searle leaves things unclear. Is he talking of real comprehension processes that a hearer must go through in order for him to understand a metaphor? Searle says that the process he describes 'need not specify a set of steps that he [a hearer] goes through consciously; instead it must provide a rational reconstruction of the inference patterns that underlie our ability to understand such metaphors' (p. 104). It is unclear from this statement if, for example, his theory is to be taken as a psychological hypothesis, subject to empirical testing. His claim that understanding the metaphorical meaning is derived from the understanding of the literal meaning of a metaphorical expression leads to the prediction that metaphors should take longer to understand than literal statements. Experiments done on this subject, such as (Harris 1976) and (Ortony *et al.* 1978), led to conflicting results, but I am not sure if Searle would take even clear cut results as bearing on his theory. Searle's statement that his proposal is 'a rational reconstruction of inference patterns' does not help much in clarifying what he means. A rational reconstruction, as far as I'm aware, is a process of finding a coherent understanding of a vague or ambiguous statement that otherwise may seem incoherent; this is not in any way what his principles are offering. Since the alternative is unclear, I will assume that Searle is talking about actual comprehension processes. As such, I will argue, his theory is inadequate.

My first claim is that the principles he describes, as a theory of comprehension, are just wrong. As I noted above, people understand all kinds of utterances, not just metaphorical ones, in various ways, using many different cues as to what a speaker may have meant by uttering an expression. The prospects of capturing any general principles of how this is done, especially by armchair reasoning, do not seem very promising and are likely to be either too narrow, or too wide to be with any content. Thus, for example, when Searle describes the first step in his strategy for metaphorical interpretation, he says that the hearer should look for a defect in the sentence uttered, or in the circumstances of its utterance.

Then he adds that defectiveness is not a necessary condition, and that just the fact that someone tends to speak metaphorically, or that what we are reading is a poetry book may tell us that an expression is meant metaphorically. If understood too narrowly, this is not enough, as one may come up with indefinitely more "techniques" for finding out that a metaphorical interpretation is in place, e.g. that of a speaker telling a hearer that he's going to utter a metaphor. If understood more broadly, this claim is vacuous, as all it says is that some thing or other will indicate that the utterance shouldn't be taken literally.

There is another, more crucial reason, for arguing that Searle shouldn't present the principled relations between the metaphorical meaning of an expression and its literal meaning as having to do with the comprehension process: this is an unnecessary complication. If these principles of comprehension work, it must be because they capture an underlying systematic relation that holds between the two kinds of meaning. But then it makes much more sense just to state these relations without being committed to the actual techniques that people employ in finding out the metaphorical meanings. Thus, for example, when Searle states principle 1, saying that 'things which are P are by definition R' (p. 107) nothing is added, except for speculation, by saying that hearers actually use this principle in the process of understanding a metaphor. Instead, he could just say that metaphors are such expressions that have a metaphorical meaning that relates to the literal meaning of the expression in the following ways: in some cases P things are by definition R things (principle 1), P things are often said or believed to be R, even though they are not in fact R (principle 2), etc. Such a way of presenting things maintains the distinction between the meaning of the metaphor itself, and things that can be said about this meaning and its relation to the expression's literal meaning. At the same time, it avoids unwarranted speculations about comprehension processes. When we claim that an expression has such and such a literal meaning, we are not by this committing ourselves to a specific interpretation process that every speaker must invariably

go through when using it. Similarly, if we do accept that there are systematic relations between an expression's metaphorical meaning and its literal meaning, we can state these relations without having to say anything about the role of this relation in the psychological processes of an individual.

Why, then, does Searle add this unnecessary complication into his explanation? I think this has to do with his trying to hold the stick from both ends, in having both systematicity and unspecificity in the meaning of new metaphorical expressions. If he said directly that metaphorical meanings can be computed reliably, by pre-established principles, then there is no point in denying that it is the metaphorical expressions that mean what they do metaphorically, not the speaker that means what he does metaphorically. By adding this "dog-leg" explanation of on-line interpretation, instead of fixed meanings, Searle manages to obscure the fact that his theory is not really making room for any novelty in the meaning of novel metaphorical expressions.

1.2.3 Searle vs. Grice

As I've pointed out earlier, Searle's account of metaphor goes along the lines of Grice's treatment of conversational implicatures. In fact, Grice explicitly mentions metaphor as a case the understanding of which can be explained by appealing to a violation of the Maxim of Quality, one of his Cooperative principles (Grice 1975, p. 53). My criticism of Searle, then, might be seen as carrying over to Grice's account of implicatures. For example, I claimed that since Searle holds that metaphorical meanings can be computed reliably, there is no point for him in denying that metaphorical expressions have metaphorical meanings, just as literal expressions have literal meaning. Now, Grice's paper was set to show how implied meanings can be computed, yet the main point that he was arguing for was that these implied meanings are of a different kind from what

the words themselves meant. As a result, he argued, such cases couldn't serve in an argument for or against the possibility and adequacy of a formal account of natural language, as they were not demonstrating divergences in the *meaning* of natural language expressions. Am I, then, by criticizing Searle, committed to an equivalent criticism of Grice's account of implicatures?

I hope I am not, as I must admit that it is absurd to hold that 'there's a garage around the corner', to use one of Grice's examples, sometimes literally means 'there is a garage around the corner that sells petrol, and is open now'. Or, to take a stronger example, when I say 'no' when someone offers me tea, the word *must* mean the same as when I use it when someone offers me coffee, rather than mean 'no, I don't want tea' at one time, and 'no, I don't want coffee' at the other. It seems necessary, then, to distinguish my claims against Searle's position on metaphor, from apparently parallel arguments against Grice's account of implicatures.

First, it should be noted that my criticism of Searle for equating metaphorical meaning and speaker's meaning, and for discussing comprehension processes, do not apply to Grice. Grice does maintain that there is a distinction between implied meaning and literal meaning, but doesn't attribute the implied meaning to the speaker, or the conversational implicatures to the comprehension process. Instead, he is describing the conversational implicatures 'as being essentially connected with certain general features of discourse' (Grice 1975, p. 45). Accordingly, Grice is not busy denying, as Searle is, that the implied meanings have to do with the objective, and publicly accessible features of the words that are used. Again, he is only trying to distinguish those systematic features from other such systematicities that should be captured by the notion of literal meaning, and which are the subject matter of semantics. In other words, he is separating semantic from pragmatic generalizations[5].

[5]I am not concerned here with judging if Grice's treatment of this problem has been successful, or if making this distinction is justified or not. I am only pointing out how his way of presenting

It may be argued that Searle's choice of words in describing the phenomenon in question was poor, but that basically the same points could have been made in Grice's terms. Would his account then have been correct? This, I agree, is the more interesting question, and I shall try to answer it. It is not a coincidence that so many discussions have been made about the notion of metaphorical meaning without directly tying them down to other cases of implied meanings. Many examples of implied meanings involve utterances of sentences that have sensible literal meaning, that could be used in the right circumstances to convey relevant information without appeal to any conversational maxims. When implied meanings are involved, finding what they are depends heavily on features of the context of the utterance. This is reflected in the nature of the conversational maxims that determine the implied meanings, that are all based on the notion of relevance, which is a clear pointer to the context. Paradigmatic cases of metaphor are different: they involve the use of utterances that are meaningless if an attempt is made to interpreted them literally. Their failure to convey information under a literal interpretation is context independent. And, most importantly, the resolution of what a metaphor might mean can be achieved (as far as it can be achieved at all), in a way that is to a large extent again context independent. Instead, it is closely related to the nature of the internal tension between the categories used within the metaphorical expression.

This difference between metaphors and cases of implied meanings, in respect to their context sensitivity, is clearly reflected in Searle's own account of metaphor. Consider Searle's suggestion for the second, most important step of the metaphorical interpretation process, which involves the PMIs. The generation of possible intended meanings, as Searle would call them, is only a function of the 'P' term in the metaphorical 'S is P' expression, and does not at all depend on any factors specific to the utterance. Or consider the third step, the process of narrowing down the possible meanings to the intended one. Searle describes

the problem is different from Searle's.

this process as one in which the possible 'R' values, generated in the second step, are matched with the 'S' term to see which combination will produce a sensible interpretation. Now, if this matching is understood as applying to 'S' as a general kind term (e.g. to a ship, in the case of 'the ship ploughed the ocean'), then this step does not depend on the context of utterance. Even if the 'S' term is a proper name, as in 'John is a pig', but the consideration taken is only that 'John' is a name of a person, and so the metaphor couldn't be attributing to him a short, curly tail, nothing depends on the context for the interpretation. Only if Searle is understood as talking about the actual *referent* of the 'S' term, does the interpretation depend on the context. But this, we saw, is both unnecessary in many cases that can be resolved by general considerations, and impossible in cases of metaphors where there just isn't a particular referent (e.g. 'light is a wave').

The only shade of context dependency we might be left with resides in the first step, where a defect in the metaphorical expression, if taken literally, is to be found. Even this defect is not always context dependent: as Searle points out, some of these defects have to do with the expression's being an 'obvious falsehood', or a case of 'semantic nonsense'. But then there are cases where the defect has to do with the expression's apparent violation of a conversational maxim, if taken literally, a clear case of context dependency. Does such a violation justify the need for a pragmatic, rather than semantic account of metaphor? Not quite. A sentence ambiguous in its literal meaning may also be defective in the same way if the wrong choice is made between its several literal meanings, and the ambiguity resolution is itself, many times, context dependent. This shows that this type of context dependency in itself is not a sufficient reason for insisting on a pragmatic treatment of metaphor. Conversely, the fact that so much of the interpretation is context independent shows how different metaphor is from the typical cases Grice was considering, and suggests that a semantic rather than pragmatic account is to be preferred.

To summarize the comparison between Searle and Grice: first, my criticism of Searle need not carry over to Grice, in so far as it is aimed at Searle's specific formulation of the problem at hand. Second, metaphor is very different from some of the cases that Grice had discussed, and for which I wouldn't want to dispute that a pragmatic account is adequate. In that respect it opens the possibility (though not the necessity) of treating metaphor differently. In particular, *if* Searle is right and the metaphorical meaning of expressions is systematically related to their literal meaning, and *if* he is right about the fact that this relation hardly depends on the context of utterance, then I see no reason why he should claim that metaphorical meaning is not a proper part of the meaning of metaphorical expressions that should be accounted for by a semantic theory. At this stage, however, I am not committing myself to any position as to what exactly a semantic theory should cover, or to the nature of the systematicities in metaphorical meaning.

1.2.4 Are metaphors systematic?

Until now I have gone along with Searle's view that metaphors are systematic in some way, i.e. that there is a systematic relation between the metaphorical meaning of an expression and its literal meaning. I only pointed out that this view is inconsistent with his attributing the metaphorical meaning to the speaker's intentions, and that assuming such systematicity made the appeal to the comprehension processes redundant in the explanation of metaphorical meaning. I also said that accepting the idea of systematicity left the other aspect of metaphorical communication, its novelty, unaccounted for. Cooper, however, doesn't accept so readily the claim that metaphors are systematic. He complains that Searle's choice of examples, all of which are established metaphors of the subject-predicate form, lends his claims about systematicity unjustified plausibility. Using such examples Searle relies on our ability to easily recog-

nize paraphrases that are associated with these metaphors, and thus boosts our intuitions about there being something standard and regular about metaphorical interpretation. If Searle were to discuss instead cases of fresh metaphors, Cooper claims, the systematicity and ease of interpretation wouldn't be so obvious. Cooper's book contains many examples of such unestablished metaphors that help make this point, such as Eliot's 'I will show you fear in a handful of dust' and Nietzsche's 'truth is a woman'.

Searle, however, had anticipated this sort of criticism, and as I mentioned above (p. 9) he replies to it by saying that the established metaphors are not only acceptable as examples because they were fresh metaphors at one stage, but that they are of particular interest because their continual use indicates that 'they satisfy some semantic need'. Is this a satisfactory answer to Cooper's criticism? In one way it is, in another way it isn't.

I believe that Cooper was right in claiming that many cases of metaphorical expressions are such that it is far from obvious if, and how, they can be paraphrased. Even if an adequate paraphrase is found, it is probably still impossible to find in those cases, in a way that is not ad-hoc, a systematic relation between the paraphrased metaphorical meaning and the literal meaning of the expression used. Nevertheless, I think that Searle has touched upon an important issue when saying that established metaphors are of special interest. They are interesting precisely because they reveal the systematicity that Searle has described, and this is probably part of the reason that enabled them to become established. The problem with Searle's choice of examples, then, is not in the choice itself but with what his examples are supposed to exemplify. The systematicity observed in established metaphors may not exemplify a systematicity that is characteristic of *all* metaphors, but may just hold in cases of metaphors that tend to become established.

What I am suggesting—and this suggestion will be developed and explained in later chapters—is another way of classifying metaphors. Except for the commonly accepted classification of metaphors into established, or 'dead' metaphors vs. fresh metaphors, I maintain that they can also be divided into one-off metaphors vs. *establishable* ones. By establishable metaphors I mean such metaphors that already on their first appearance have certain qualities that enable them to become established. These qualities have to do, among other things, with the systematic relation between how they are interpreted, and the literal meaning of those expressions. This division will give us part of the answer to our dilemma of how metaphors are regular and irregular in their meanings at the same time. The one-off cases are those that reinforce our intuitions that metaphors are irregular in a way that defies any generalizations made about metaphorical meaning. The establishable metaphors are those that can be subject to such generalizations. I said, however, that this is only part of the answer to the dilemma. As I will later argue, the establishable metaphors themselves are systematic only to a certain extent: the meanings that they may acquire as they become established is constrained, but not fixed. Claims about systematicity in the metaphorical meaning, such as those Searle has offered, should be understood as spelling out the constraints to the acquired meaning of metaphors, rather than ways of exactly computing what these meanings may be. I realize that my claims here are quite vague at the moment, but their full explanation will have to wait for later presentation. At this stage, I only wish to point out the possibility of reconciling Cooper's criticism with Searle's intuitions that metaphors are systematic, by emphasizing the difference between one-off metaphors and establishable ones.

It is worth noting that Cooper himself, discussing a different matter, recognizes a systematicity in established metaphors. He approvingly mentions Lakoff and Johnson's book 'Metaphors We Live By' (1980) which centers around the idea that our metaphorical talk is highly systematic.

> 'it is the exception, rather than the rule, for established metaphorical expressions to have become established singly. More typically, it is an expression along with many other related expressions which, *en bloc*, as it were, develop a new usage outside of the parent domain'
>
> (Cooper 1986, pp. 130–1)

The systematicity he is talking about here is not the one Searle was talking about[6]. It is not the systematic relation between an expression's metaphorical meaning and its literal meaning that is discussed, but instead the phenomenon of metaphors' appearance in batches of related expressions. One of the examples of such a system of metaphors, taken from Lakoff and Johnson, is that which centers around the basic 'time is money' metaphor. Along with it we get metaphors such as 'you are wasting my time', 'I don't have time to give you', 'how do you spend your time?', 'I've invested a lot of time in her', etc. But even though Cooper is not talking about the same kind of systematicity, his discussion points out again that generalizations can be made, and systematicity found in established metaphors that cannot be made about metaphors in general. It would be misplaced to argue that metaphors do not display the systematicity that Lakoff and Johnson claim they do by bringing up examples such as those which Cooper used against Searle. 'Truth is a woman' doesn't have other related metaphors because it isn't an established metaphor. Similarly, if we accept that Searle's generalizations hold only for metaphors that become established, such counter-examples cease to have a force against his claims.

If my suggestion that Searle's account applies only to establishable metaphors is correct, then he hasn't provided us with a full answer to his basic question, 'how do metaphors work?'. At best, he has explained how some metaphors work. As I have argued, however, this question was a bad one to begin with because it leads to giving inadequate, quasi-psychological answers. If this question can be answered at all, the answer should better be backed up with some

[6] I will later argue that though these systematicities are different, they are not unrelated.

hard psychological data. I don't think, therefore, that Searle even explained how some metaphors work. But all this doesn't mean that his theory is worthless—It just means that he was answering somewhat different questions from the one he had posed. What I think he has shown us is that there are systematic relations, that can be stated along the lines his PMIs suggest, between the literal meaning of expressions and the additional meanings they may acquire as they become established. This need not be treated as a first step in a theory that can be further developed to cover all cases, as it is probably the case that not all metaphors can have generalizations made about their meanings. This, we shall see, is a view that lead some theorists (such as Cooper and Davidson) to claim that metaphors just do not have a meaning.

1.2.5 Conclusion

Searle has attempted to capture in his theory the two conflicting aspects of the metaphorical use of linguistic expressions. Since presumably fresh metaphors involve a use of words in a non-standard way, he attempted to locate the metaphorical meaning in the speaker's intentions for communication. He did not take, however, the consequences of such a move into full account. In the next chapter we shall see what grave consequences such a move has, as we follow Davidson's claim that literal meaning should be accounted for in terms of the speaker's intentions. Appreciating the fact that such intentions cannot be restricted, Davidson concludes that all systematicity found in the relation between words and their meanings can only be stated in ad-hoc terms. As a result he will claim that 'there is no such thing as language', where a language is understood as predetermined and shared set of rules for interpretation.

Chapter 2

Davidson on Metaphor

2.1 Metaphors and Malapropisms: Davidson on the Limits of the Literal

2.1.1 Introduction

Donald Davidson discussed, in two separate papers, two irregularities in linguistic communication: metaphors (Davidson 1984) and malapropisms (Davidson 1986)[1]. Both papers were put under the same heading—'Limits of the Literal', in two collections of his works. Davidson himself did not discuss the relation between these phenomena and his accounts of them, and I would like to fill in this gap. I find it important, first, because a discussion of this relation gives a better understanding of both accounts, and of what he takes to be the literal and its limits. Secondly, some apparent similarities between Davidson's discussion of the two phenomena may mislead a reader into thinking that they are treated in the same manner. Richard Rorty, for one, was mislead. In a footnote to his paper 'Unfamiliar Noises' he says: 'See Davidson's 'A Nice Derangement of Epitaphs' for a parallel between metaphors and malapropisms' (Rorty 1987,

[1] All page references to Davidson in this section, unless otherwise stated, are from this paper.

p. 290, footnote 18). To me it seems that appreciating the *contrast* between the two accounts is essential to the understanding of Davidson's views on these issues, and so I will show how those treatments differ.

The thrust of Davidson's argument in 'A Nice Derangement of Epitaphs' goes against the view that what a sentence means (i.e. its literal meaning), in contrast with what a speaker means, depends upon the conventional or regular way that it is interpreted by a community. The argument against this view emerges from the observation that too many utterances in actual communication situations are irregular, without communication being impaired. Examples of such irregularities are malapropisms, slips of tongue, new idiolects and more. Davidson describes a situation of successful communication as one where a speaker intends his utterance to be interpreted as true if and only if certain truth conditions hold, and his hearer interprets him just in the way that was intended, i.e. recognizes what the relevant truth conditions are. The problem that arises is that a single standardized semantic theory, based on an abstraction over regularities in communication, will not give the truth conditions of many of the sentences that are uttered and understood. Specifying the truth conditions of every sentence in language is one of the minimal requirements of any adequate semantic theory, Davidson thinks, and so such a general theory fails. Moreover, it is not the case that a multitude of such semantic theories, as detailed as we would want to have them, would do the job. This is because many of the irregularities mentioned above are created at random, and so cannot, in principle, be accounted for in advance of their being uttered.

Davidson concludes that the notion of literal meaning, which is the subject matter of semantics, has to be accounted for in a way that will not depend on regularities in communication. What he offers as an alternative, together with what he takes to be an adequate description of the communication process, will be presented in the first part of this section.

In the argument just presented, Davidson does not deny that there are standard meanings. These are given by 'what a good dictionary would say, or what would be found by polling a pod of experts whose taste or training I trust' (p. 434). Neither does Davidson think there is no such thing as literal meaning. He only denies the identity between the two, and thus commits himself to provide an alternative analysis of that very notion of literal meaning. This denial is based, we should notice, on the assumption that literal meaning is the subject matter of successful communication, i.e. that (at least an important part of) what is communicated is, by definition, the literal meaning of what is uttered. For some, this may seem a plausible assumption, for others the source of a defective argument. Before launching an attack on this assumption, however, one should bear in mind that a notion of meaning that does not make it an essential part of the communication process may be somewhat uninteresting. This, in fact, is what Davidson thinks is the fate of the notion of standard meaning (see p. 434).

Explicating literal meaning in terms other than regularities seems somewhat arbitrary, as it goes against the standard definition of the literal ('Literal: taking words in their *usual* or *primary* sense and applying the *ordinary* rules of grammar, without mysticism or allegory or metaphor'. The OED, my italics). To avoid this problem Davidson introduces a technical term, 'first meaning', as a replacement. However, he takes first meaning to be the subject matter of semantic theories, and thus gives it the role that other philosophers give to literal meaning. He uses the two notions interchangeably throughout his paper, and so will I. The reader is therefore warned not to be confused by equating 'sentence meaning' and 'literal meaning' on the one hand with 'standard' or 'regular meaning' on the other. Justifiably or not, these are contrasted, rather than identified, in Davidson's account.

The second part of this section will deal with the application of Davidson's

account of literal meaning to what created the problem in the first place— malapropisms and their kin. After describing what Davidson takes to be an adequate account of malapropisms, I will turn to distinguish it from his account of metaphor. The point to be emphasized by this contrast is that again, contrary to the common view (as we have seen above, expressed in the dictionary definition), Davidson thinks that metaphors are different from the literal *not* because one is regularized and the other is not. Instead, it is that what other people call metaphorical meaning is of a different kind than literal meaning, and has nothing to do with semantics. This point is not new to Davidson's account of metaphor, but gives it a different stress. I will try to show how the ideas expressed in 'A Nice Derangement' reflect on Davidson's earlier discussion on metaphor. Through this discussion I will present Davidson's account of metaphor, that will be discussed in further detail in the second half of this chapter.

2.1.2 Literal meaning and communication

To begin with, let us consider Davidson's characterization of literal meaning, or as he prefers to call it, 'first meaning'. 'The concept applies', Davidson says, 'to words and sentences as uttered by a particular speaker on a particular occasion' (p. 434). First meaning is distinguished by the order of the speaker's intentions: when uttering a sentence S, the first intention of a speaker is that his sentence will be interpreted by his audience as true if and only if p (where p is a specification of some truth conditions). A further intention, in many cases, is to express a belief that p holds; next he might intend to convince his hearer that p, by way of having expressed his own belief that p; other intentions may still follow. The first intention in that sequence—that of having H interpret S in a certain way, specifies the first meaning of S. The meaning specified by this intention are the truth conditions that must hold for that sentence to be true.

For example: U utters the sentence 'clowns are funny' and his first intention is that H will interpret his words as true if and only if clowns are funny. Next, U intends to express a belief that clowns are funny. Then, he intends H to come to believe that clowns are funny, knowing that H will take his word for it (U's first two intentions need to include H's recognition of these intentions as such, in the Gricean manner, so that H will take U as actually saying that p). A further intention may be to cause H to buy tickets to the circus, with the ultimate intention of (what else...) making some money (U owns the circus). Here, the first meaning of 'clowns are funny' is that clowns are funny. Obviously, all the higher level intentions can be attempted by means other than uttering S: U can laugh very loudly in the presence of H, when seeing a clown; or, he may show H a picture of a clown with people around him, laughing, etc. Conversely, U may utter S with the same first intention, yet he may have different higher level intentions: his intentions may not be to express belief that p, but rather to irritate H who hates clowns. In both cases though, S had the same first meaning, since U had the same first intention.

This way of characterizing first meaning clarifies the distinction between sentence, or literal, meaning and speaker's meaning: the first intention specifies the literal meaning of an utterance; the higher level intentions are what the speaker means, that is, what he intends to achieve by uttering that sentence (e.g. expressing beliefs, convincing, pleasing, scaring or any other purpose).

The concept of first meaning applies to any sign that has non-natural meaning, not just to linguistic signs. Linguistic meaning is a special case of first meaning, distinguished by (at least) two further conditions that must hold: (i) that first meaning is systematic; i.e. there must be systematic relations between the meaning of utterances. (ii) First meanings are shared—successful communication depends on sharing the system that was mentioned in (i) (see p. 436). The system should be thought of 'as a machine which, when fed an arbitrary

utterance ... produces an interpretation' (p. 437). This system is modeled by a semantic theory, that has a finite base and is recursive.

The intention that specifies first meaning in a case of linguistic meaning is more complex than in the general case: it is that the speaker's utterance will be taken as true if certain conditions hold, *by virtue of the semantic properties of its parts*, where these properties are given by a semantic system. An analogy may help explain these requirements: a move of a piece on a board is given it's meaning as a winning move of a game by the system of rules that define that game. If one player moves a piece with the intention of winning a game, but thinks he is playing chess, while the other understands it to be the winning move of a different game (i.e. a game with a different set of rules), all we get is a confusion that was caused by moving a piece of wood on a board. Only if they share the system of rules of the game, i.e. one player intends his move to be a move of chess, and the other takes it as such, will the move be one of a chess game between the two. Analogously, only if a speaker intends a sound he makes to have a certain meaning by virtue of a set of interpretation rules that applies to that sound, and his hearer shares with him the same set of rules, do this speaker and hearer communicate linguistically.

The process of linguistic communication is described by Davidson as follows, starting from the hearer's perspective: before conversation begins, the hearer has a semantic theory that he believes will be adequate for the interpretation of a specific speaker's utterances. Davidson calls this the hearer's *prior theory.* If the hearer knows very little about the speaker, his prior theory will be based on a generalization that he makes over speakers of, say, English. In our above analogy with games, a player may play, say, poker with a partner he had never met before. Poker has many versions and specific rules vary from one place to another. If our player knows nothing about his partner, he may assume that they are going to play the version which he believes is the most common. The

set of rules for this version is his prior poker theory (PPT) for that partner.

Back to language: in every case, the hearer will have some adjustments made in his theory, to account for what he expects to be the speaker's idiolect; these may be an approximation based on stereotypes that he has on how people of a certain type speak. Thus, a hearer may prepare differently to interpret a taxi driver than he would prepare to interpret a politician: he will assume that one set of rules will best serve him in interpreting the first, and a somewhat different set will be better in interpreting the other. The more a hearer knows about a speaker, he can make more and better such adjustments. In poker terms: If our player is playing in a bar with a cowboy, he may use a different PPT than if he is playing with the madam of that house, since he knows that ladies play with somewhat different rules of poker. The more he knows about his partner—that he usually plays at that specific bar, and therefore is likely to be familiar with the specific rules there, etc. —the better are the chances that he will pick the right PPT.

Conversation again: after hearing an utterance from a speaker, the hearer may realize that further adjustments are needed in order to get the most reasonable interpretation of what he has just heard. He will then create a new, ad-hoc, semantic theory, and use it for the interpretation of that utterance. The theory that he actually uses is called the *passing theory*. The hearer may then decide that the deviation he had just encountered from his prior theory is likely to recur, and accordingly he will change his prior theory to be the same as the passing theory he has just used. The new prior theory is what he will be equipped with when he listens to that speaker's next utterance. Alternatively, he may think that the deviation was not chronic but just, say, a slip of tongue. He may then refrain from making changes in the prior theory, and will keep the original one for interpreting the following utterance. The prior theory, in this sense, can be viewed as the hearer's long-term strategy for interpreting the

speaker[2]. As conversation continues, the hearer may keep changing the passing theory, and maybe the prior theory, as needed.

Well, here our poker analogy may break down—the players should better decide explicitly what rules they are going to use, or shooting may begin. But if we imagine a more civilized part of the world, and a game on friendly terms, we can assume that during the game our player will see by certain moves his partner makes, what version exactly he is playing. The new sets of rules that he will come up with during different stages of their game are his passing poker theories, and these will also be the adjustments he will make to the prior poker theory he uses with that partner. In conversation the situation is different in two respects: first, natural languages are so complex that there is no chance of specifying in advance all the rules that are going to apply (even if people knew explicitly what these rules are, which they don't), and these are to be discovered as conversation develops. Second, as Davidson argues in his paper, in language there need not be any appeal to a predetermined set of rules; competent hearers have the resources to come up with sets of rules (theories) that fit the data without having to learn all possible theories in advance.

On the speaker's side we have a similar situation to that described for the hearer. His prior theory will be the one that he believes his hearer is going to use for interpreting him. He may intend the hearer to use that theory, or he may want him to use a different theory for interpreting his utterance. The theory that he intends the hearer to use is the speaker's passing theory. Communication will be completely successful when the speaker's and hearer's passing theories converge; or, in other words, when the speaker is in fact interpreted as he intended. When communication is thus successful, the first meaning of a sentence uttered is what was specified by the speaker's first intention[3].

[2]See (Dummett 1986) for a development of this point.

[3]It is not entirely clear if Davidson demands that for words to have a certain meaning, as intended by the speaker, successful communication has to actually take place. He might be

As we can see, this account suits Davidson's view that the concept of literal meaning should apply 'to words and sentences as uttered by a particular speaker on a particular occasion'. The account completely ignores normative notions such as 'standard use', 'correct use' or 'mistaken use'. Davidson plays down the importance of these notions in the discussion of language and communication: '...error or mistake of this kind, with its associated notion of correct usage, is not philosophically interesting. We want a deeper notion of what words, when spoken in context, mean' (p. 434). The reason that these notions are said to be philosophically uninteresting is that they are based upon regularities abstracted from a community's use of language. As explained above, these regularities cannot be the basis of a satisfactory semantic theory, since that theory will not give the truth conditions for all sentences that need to be accounted for.

Davidson concludes from his way of viewing communication, that 'there is no such thing as a language, not if a language is anything like what many philosophers and linguists have supposed' (p. 446). Instead, there are many idiolects that are constantly changing to accommodate for different conversation situations. Linguistic competence, instead of being the mastery of one particular language, is the capability of adjusting passing theories, to achieve maximal convergence between speaker and hearer.

2.1.3 Metaphors and malapropisms

After familiarizing ourselves with Davidson's characterization of literal meaning and the communication process, let us see how it applies to malapropisms, the problematic case that was claimed to make other accounts inadequate. When Mrs. Malaprop said 'a nice derangement of epitaphs', her first intention was

content with there being sufficient resources for an interpreter to get at the right theory, without insisting on success. While this is an important point in understanding Davidson's account, it is not crucial for our discussion here. We will assume that he does require actual success in communication, though this might not be necessary.

to be interpreted as saying (what in standard English is expressed by) a nice arrangement of epithets. If she was actually understood as she intended—say, by someone that recognized her mistake, but nevertheless knew what she meant, then her words literally meant, according to Davidson, a nice arrangement of epithets. That would be the correct translation into standard English from the language that she had spoken to her hearer.

This was a case where understanding was achieved though one side was mistaken about the standard meaning of his words. The same thing could happen deliberately, when someone changes one word for another, say to make a pun. Again, it could happen because both speaker and hearer make the same mistake, and thus understand each other without either of them being aware that a mistake has been made. In all these cases, Davidson says, what actually was communicated defines what the words or sentence uttered literally meant on that occasion; and what picks out what that meaning was, is the speaker's first intention. It is not the case, however, that 'anything goes', i.e. that words can mean anything their speaker intends them to mean. They get their meaning only if communication is successful (see footnote 3), and when it is, it no longer matters what they regularly, or standardly, mean.

It may seem that Davidson, by characterizing literal meaning in terms of a speaker's intention, have come very close to loosing the distinction between what a speaker means, and what his words, or sentence, mean. Yet Davidson insists 'that nothing should be allowed to obliterate or even blur the distinction between speaker's meaning and literal meaning' (p. 434). The way the distinction is maintained in his account is by its appeal to the order of intentions, together with the demand that communication is in fact successful. It is the first intention that gives a sentence its literal meaning, and this is what the semantic theory should account for. The higher levels of intention are what is called speaker's meaning, and these are of no concern to the semanticist. We will next look at

how this distinction is applied, in particular for distinguishing metaphors and malapropisms.

Davidson demonstrates the distinction between speaker's meaning and literal meaning in discussing an exchange between K. Donnellan and A. Mackay. Donnellan has distinguished between a referential and an attributive use of definite descriptions. The distinction is illustrated by the following: Jones says 'Smith's murderer is insane', having in mind a certain man that he believes has murdered Smith, and whom he thinks is insane. Donnellan claims that the definite description 'Smith's murderer' was used referentially, and that Jones has said something true if the man he had in mind was insane, no matter if he actually was Smith's murderer or not. On the other hand, Jones may have used this sentence attributively, meaning something like 'Smith's murderer, whoever he may be, is insane'. In this case the truth of what he said will depend on the sanity of the real murderer of Smith. Mackay, in reply to Donnellan, accused him of holding an impossible theory of meaning, in which words and sentences mean whatever their speaker wants them to mean.

Davidson says that Donnellan is right in making his distinction, but that the distinction doesn't deal with word and sentence meanings at all. Instead, it has to do with the speaker's meaning. In both cases, Jones had the same first intention, that his sentence S would be interpreted in a way that referred to Smith's murderer. The difference between the two cases appears on the higher levels of intention: in the referential case, Jones had a higher level intention of expressing a belief that some particular man was insane, and he hoped to fulfill that intention by way of uttering S. In the attributive case, Jones' higher level intention was to express belief in what was expressed by S. Thus, in both cases 'Smith's murderer' has referred to Smith's murderer, and the *sentence* was true or false depending on whether Smith's murderer was in fact insane. It is just that in the referential use, *Jones* may have referred to someone else, by using

words that referred to Smith's murderer.

We have seen above that sentences with the same first meaning may serve different purposes on higher levels of intention. The referential use vs. the attributive use of definite descriptions is an example for this. In the first, as Davidson puts it, 'Jones has said something true by using a sentence that is false. This is done intentionally all the time, for example in irony or metaphor' (p. 440); in the second, the truth value of Jones' sentence and the truthfulness of what he said coincide[4]. In the example above, Jones knew exactly what his words standardly meant, and intended them to be interpreted in the standard way. He was mistaken about the facts—about who really was Smith's murderer—not about language. An utterer of a metaphor, similarly, knows very well what his words standardly mean, and normally intends them to be interpreted in a way that corresponds to the standard way. The effect that he is trying to achieve—to create an image, to suggest an analogy, or whatever, is to be explained on the higher levels of his intentions. Accounting for these higher level intentions, Davidson claims, is not something semantics should deal with.

Malapropisms, we have seen, are a different matter, according to Davidson's account. Those are cases where words are uttered with the intention of their being interpreted in a particular way, that is different from their standard interpretation. More importantly for Davidson, they are intended to be interpreted in a way that is not accommodated by the theory that the hearer has in advance of hearing the utterance, for the speaker's language. For successful communication to occur, the hearer has to make a change in his theory of meaning for that speaker's idiolect. And if the correct change will be made, then the malapropism

[4]It is not completely clear what 'saying' means in the above quote. It seems that Davidson must mean something that has to do with higher level intentions. Probably, it means that the speaker says that p if he can be taken as believing that p, and this can be done by means other than uttering a sentence which expresses p. The speaker can say something true by using a false sentence, if what he is taken to believe in is true, yet the sentence he uttered is false. This was the case with Jones' utterance. I suspect that many would go against this use of 'say', but for the purpose of understanding Davidson's point it suffices.

will get a new meaning, the one that was intended for it by the speaker and was understood by the hearer. Malapropisms, then, do have literal meaning that is non-standard, and they are a semantic phenomenon that affects interpretation.

The difference that was just pointed out between metaphors and malapropisms may seem to suggest that only malapropisms initiate changes in semantic theories, whereas metaphors should not have this effect because of their dependence on regular interpretations. This conclusion needs some qualification, because we are familiar with the phenomenon of 'dead' metaphor, i.e. metaphors that recur, and are said to acquire a new literal meaning. Davidson in his paper on metaphor acknowledges this possibility (cf. Davidson 1984, pp. 252–3) and thus admits that metaphors do initiate changes in our semantic theories. But, there is a difference between metaphors and malapropisms in the manner in which the changes are introduced, and the significance of these changes. In the case a malapropism, the change is in response to one particular utterance, and immediately affects the passing theory that is used to interpret that utterance. As was explained earlier, the hearer may then consider, after making the change to the passing theory, whether it is wise to incorporate the same changes into his long-term, prior theory. The change in the passing theory, in contrast, is forced, so to speak, on the hearer. If it is not made, the utterance will just not make sense, and the speaker will probably not achieve the higher level intentions he had. If, for example, Mrs. Malaprop meant her utterance of 'a nice derangement of epitaphs' as a compliment, but one did not make the necessary adjustments to his theory for interpreting her, he would not be likely to understand that she was trying to give him a compliment.

With metaphor, it is a different kind of change. One understands a metaphor, or rather the higher level intentions of its utterer, by interpreting the speaker as he usually would. But, he may notice over time, that this speaker is using that metaphor again and again, and that it is used to convey some particular

proposition. In this case the hearer will probably make a change directly into his long-term, prior theory. This will not be a response to some particular utterance of the speaker, but rather to the observation of some recurring pattern that the speaker is exhibiting. This will probably result, as it actually seems to be the case with 'dead' metaphors, in blurring the connection between this new lexical item and its origin as a metaphor. The change in the case of metaphor, thus, is not essential to understanding the speaker. Quite the contrary—usually there is a point to one's making a metaphor, and this point will be lost if one takes the metaphorical utterance as an irregularity that calls for some change in interpretation. The later change is made only as a strategic change, when the point had already been made again and again to the stage of being dull. The hearer can now use a short-cut in understanding what the speaker is trying to get at, without loosing anything. The change is made only when there is no longer a need for 'the primary or original meanings of words [to] remain active in their metaphorical setting' (Davidson 1984, p. 249).

There are some parallels between malapropisms and metaphors, as Rorty has put it, but these are only superficial: most metaphors are false sentences; many malapropisms, if wrongly interpreted in a standard way, are also false sentences. But there is an essential difference between the two cases: metaphors are *intended* to be interpreted as false sentences. Malapropisms, on the other hand, may be taken as false sentences only if interpreted along some standard lines. But these, as Davidson explained, do not determine the sentence's literal meaning; if interpreted as intended, malapropisms will in most cases be true sentences (this does not have to be the case—a malapropism may occur in the utterance of irony or metaphor, so that the utterance may be intended to be interpreted, non-standardly, as a false sentence).

An even closer parallel can be found in what Davidson says about the understanding of malapropisms and metaphors. On metaphor he says: 'understanding

a metaphor is as much a creative endeavor as making a metaphor, and as little guided by rules'; and, 'there are no instructions for devising metaphors; there is no manual for determining what a metaphor 'means' or 'says'' (Davidson 1984, p. 245). For malapropisms, their interpretation depends on creating what Davidson calls a passing theory. But, 'there are no rules for arriving at passing theories'; '...it is derived by wit, luck, and wisdom'; 'there is no chance of regularizing, or teaching, this process than there is of regularizing or teaching the process of creating new theories...'; '[there is] no portable interpreting machine set to grind out the meaning of an arbitrary utterance' (pp. 445–6).

It is essential to the understanding of Davidson's position on metaphor and on malapropisms, to see that while the words he uses for discussing both are similar, he is talking about different things. The meaning of the malapropisms that isn't 'ground out by a machine' is their *literal meaning* (i.e. sentence, or word, meaning). The 'meaning' (and Davidson uses single quotes) of metaphor that isn't 'determined by a manual' is the *speaker's meaning*. The creativity that is needed to create passing theories is specific to the process of understanding natural language, and is what we call linguistic competence. The creativity that is involved in creating and understanding metaphor is as general as the number of uses language can serve—that is, everything we can talk about.

The fact that a previously held semantic theory is not enough to account either for the understanding of non-literal expressions, or for the interpretation of malapropisms, doesn't show that what does account for understanding in both cases is the same, or is parallel in any interesting way. Quite the contrary: all it shows is that using an utterance in a non-standard way does not suffice to distinguish it as a non-literal expression. In the case of metaphor, the use is non-standard in that people mostly intend to express belief in a sentence they have uttered, but with metaphors they intend to achieve something else. In the case of malapropisms, the use of words is non-standard in that a speaker intends

them to be interpreted in a way that is different from the standard interpretation (as far as there is one).

If we look back at Davidson's earlier paper, 'What Metaphors Mean', in the light of the comparison with his account of malapropisms, we see that while his main claim in that paper remains unchanged, there is a difference in stress. His main claim in the paper on metaphor was that metaphors are not a semantic phenomenon, and are a matter of language use rather than meaning. In our discussion here we have seen a similar claim, that metaphors are to be accounted for by appeal to the speaker's meaning, rather than their literal meaning. 'Speaker's meaning' and 'language use' seem both to refer to the higher levels of intention in communication. In this sense, the main claim hasn't changed.

A central claim Davidson makes is that 'a metaphor doesn't say anything beyond its literal meaning (nor does its maker say anything, in using the metaphor, beyond the literal)' (Davidson 1984, p. 262). The first half of this quote is the (by now) familiar claim that metaphors, on the level of (first) meaning, do not call for any special interpretation, but just mean what would normally be taken as their interpretation. The second half, though, may seem surprising, and in contradiction with what was quoted earlier from 'A Nice Derangement': 'Jones has said something true by using a sentence that is false. This is done intentionally all the time, for example in irony or metaphor' (p. 440). Here Davidson seems to maintain that with metaphor speakers *say* true things by using false sentences, whereas in the original paper on metaphor he claimed that not even the speaker says anything beyond the literal, usually false, meaning of the sentence. Can this apparent conflict be resolved?

First, let us see exactly what Davidson means when he claims that the speaker is not saying anything beyond the literal. If we look towards the end of the paper on metaphor, we see that Davidson denies that a metaphor has 'a

specific cognitive content' (Davidson 1984, p. 262). Instead, he says, 'there is no limit to what a metaphor calls to our attention, and much of what we are caused to notice is not propositional in character' (*ibid*, p. 263). In our new terms, what Davidson is saying is that with metaphor, even the higher level intentions do not have a specific content, i.e. the speaker is not trying to relate in any way to any specific proposition. It is in this sense that even the maker of a metaphor is not saying anything.

Our problem can be resolved, by reading the quote from 'A Nice Derangement' a little differently. It is not saying that metaphor is like the referential case in that the speaker says *propositionally* true things by using a false sentence. All it says is that metaphors, like the referential case, use false sentences to achieve something different than expressing belief in what was uttered. The cases are similar in being properly explained on higher levels of intention than the first, meaning level. As Davidson puts it, 'Once we understand a metaphor we can call what we grasp the 'metaphorical truth'' (*ibid*, p. 247) but this is truth in a different, loose sense, from the notion of truth we use in doing semantics. It is in this sense only, that we can say that the speaker of a metaphor was, like Jones, 'saying something true by using a false sentence'. But we should notice that in contrast with metaphor, Jones did 'say' something with a propositional content, on the higher level of intention: he said that a certain man he had in mind was insane. We must conclude, then, that while Davidson thinks metaphors do not have a specific cognitive or propositional content, this fact cannot serve as the basis for distinguishing, in general, literal from non-literal expressions. Jones, we saw, expressed a specific proposition non-literally.

Davidson chose to open his paper on metaphor by pointing out, as I quoted above, that understanding metaphor is 'a creative endeavor', not guided by rules. After having looked at his view on malapropisms, we see that this also does not distinguish metaphors, as non-literal expressions, from literal expressions;

understanding the latter too, in many cases, cannot be explained by appeal to an existing system of rules. This cannot be taken, therefore, as a point in Davidson's argument for the distinction.

The relevance of this claim should be understood, instead, in light of his saying that metaphors do not have a specific propositional content: anything a metaphor calls to our attention, propositional or not, is a creative addition to what was really there, on the semantic or any other level of intention. As I explained earlier, it is a creativity of a different kind than the creativity needed for coming up with a new theory that explains new data. The latter creativity is the one used in organizing new semantic facts, i.e. for understanding what certain words or a sentence literally meant; it is what is needed for the discovery of the rules of a semantic theory that is assumed to exist as the speaker's passing theory. In the case of metaphor, in contrast, no such underlying system is assumed to exist, and no meanings are expected to be revealed. Instead, understanding a metaphor is the process of further creative development of what has been said and already fully understood on the semantic level.

Pointing out the creativity needed for understanding metaphor, thus, should be taken as part of Davidson's claim on the non-cognitive, non-propositional nature of 'metaphorical meaning'. But this, we have seen, is a peculiarity of metaphor, not of non-literal expressions in general. It is not what distinguishes the literal from the non-literal, as the non-literal may, in principal, have specific propositional content.

What does distinguish literal from non-literal expressions, in Davidson's account? It is not that the first always have regular, rule-guided interpretations that the latter do not have; neither is it the case that the first have specific propositional content that the latter always lack. We see now that Davidson did not attempt to fully answer this question in his paper on metaphor. The

account he had given there was specific to metaphor in saying they do not have any particular propositional content; it cannot therefore be generalized to all non-literal expressions. The question of what are the limits of the literal was dealt with only in his later paper, 'A Nice Derangement'. The answer given there was that the literal is defined by a speaker's first intention, of how he is to be interpreted, in a successful case of communication.

The question remains, however, how do we recognize a speaker's first intention? Indications as to what the answer is can be found in the paper on metaphor: if we have a reason to believe that the speaker would be rational in using certain words, under our current interpretation, to achieve what we take to be his higher level intentions, we do not need to make changes to our theory of interpretation. In other words, we may assume that he intended to be interpreted in that way. With metaphor, in particular, we may even have a reason to believe that *only* by depending on the existing meaning of words, can the speaker hope to achieve those higher level intentions. That is because the point the speaker is trying to make is such that it cannot be paraphrased, i.e. it is not one particular proposition that can be expressed by a different sentence that will make just that point. Conversely, the interpretation we arrive at in using our prior theory might make the speaker seem irrational, because his words just do not make sense to us. The principle of charity guides us then to make the necessary adjustments to our prior theory by providing a passing theory that will correct the situation.

To summarize, let us state the relations between metaphors, malapropisms, literal meaning, speaker's meaning, and regularities in communication, according to Davidson's views as expressed in the two papers:

- Literal meaning is not a matter of regularities: it is defined only by the speaker's first intention.

EX BIBL. UNIV. EDINBURGEN.

- As far as literal meaning is concerned, metaphors are not irregularities. Metaphorical sentences, therefore, should not initiate immediate changes in the way of interpreting a speaker.

- As far as literal meaning is concerned, malapropisms are irregularities. They do cause immediate changes in interpreting a speaker.

- The speaker's meaning, in cases of metaphor, is usually irregular: the speaker is using a false sentence to achieve intentions other than expressing belief in what he literally meant. This leaves open the possibility that his higher level intentions do not relate to any particular propositional content.

- Finally, the speaker's meaning, in cases of malapropisms, is usually regular: he does intend to to express a belief in what he literally meant, among his other intentions. This, however, is not necessary, as malapropisms are only defined relative to standard meaning. A speaker may have, in principal, any higher level intentions to accompany his meaning intentions.

2.2 Discussion

In the exposition above, we have seen a clear demonstration of what an account that attributes meaning to words and sentences according to their speaker's intentions should look like. There are no pre-determined constraints on an expression's meaning; the theorizing follows the intentions, rather than the intentions following the theory, so to speak. Searle has got this the other way around. The exposition discussed in great detail the phenomena of malapropisms and their kin, which I find important in itself, but also for clarifying the issue of metaphor, which is the main concern of this thesis. In the following, I will go back to concentrating on Davidson's account of metaphor[5].

[5]All page references in the rest of this chapter are to (Davidson 1984), unless otherwise stated.

2.2.1 Davidson's view on systematicity in metaphor

Our basic theme, to reiterate, is the tension in the relation between the literal meaning that words have, and their metaphorical meaning: is this relation systematic? Davidson insists, for many different reasons, that whatever a metaphor causes, only its literal interpretation captures its meaning. There is no such thing, therefore, as metaphorical meaning. I will try to go along with Davidson on this point, and restate the basic question in terms that will agree with his formulation. Is there a systematic relation between what a metaphor literally means, and the effects it has on hearers? What Davidson's answer to this question may be is not straightforwardly clear. He says that his claim that neither the metaphor nor its utterer mean anything beyond the literal 'is not, of course, to deny that a metaphor has a point' (p. 246). Presumably, then, there might be a systematic relation between a metaphor and 'its point'. But Davidson insists that this point cannot be equated with any specific fact, or cognitive content. That is, with something that can be stated by a proposition. He says for example: 'What I deny is that metaphor does its work by having a special meaning, a *specific* cognitive content' (p. 262, my italics). And 'Joke or dream or metaphor can, like a picture or a bump on the head, make us appreciate some fact—but not by standing for, or expressing, the fact' (p. 262). In that, we have seen, it is distinguished from malapropisms and their kin, where there is such a fact or proposition to be sought for.

I find it difficult to understand, and sensibly discuss, the obscure notion of 'having a point', and so I will restate our problem once again in a way that will enable a relatively clear discussion: is there a systematic relation between what a metaphor literally means, and those effects it has on hearers with a content that can be stated in terms of a proposition or a fact? Two things should be noted about this formulation. First, it is not asking whether the metaphor is paraphrasable in any existing language. I allow that there are limits to the

expressive powers of any given language, and thus not every proposition or fact can be linguistically expressed. The question thus concerns a systematic relation between a metaphorical utterance and a proposition or a fact, not a relation with another sentence, which is a linguistic entity, that might express the same fact. Second, this formulation is not a retreat to the first one. It is not asking about metaphorical meaning, but about a propositional content that may or may not be awarded the Victoria Cross of meaning. Davidson himself stresses the point that his argument is not dependent on any particular conception of meaning such that it could be disposed of by 'shrugging it off as no more than an insistence or restraint in using the word 'meaning'' (p. 262).

To this question, in its third formulation, Davidson's answer is clearly negative. There is no unique fact or proposition that is systematically related to a metaphor, apart from the one expressed by the metaphor's literal interpretation.

2.2.2 The basic dilemma again

Think now of the metaphor that Davidson begins his paper with: 'metaphor is the dreamwork of language'. Suppose someone reads it and is caused by it to think 'metaphors use symbols', because he knows that dreamwork involves the interpretation of symbols. Someone else also reads the same line, and is caused by it to think 'John will be late for school' because the mentioning of dreamwork reminded him that John was still asleep, etc. Both these thoughts were, by assumption, caused by the metaphor. If we add to this assumption that both thoughts were true, we may then call both these thoughts a metaphorical truth, as according to Davidson 'there is no reason ... not to say these visions, thoughts, and feelings inspired by the metaphor are true or false' (p. 257).

Something is clearly inadequate in this account. The first thought, 'metaphors use symbols', is properly related to the meaning of the words that were

uttered (or written, in this case) in a way that has to do with the notion of metaphor; the second thought isn't so related. Capturing this difference must be part of an adequate account of metaphor. Presumably, the difference must be stated in terms of some kind of systematicity that holds in the first case and doesn't in the latter[6].

An account such as Searle's, that argues for systematicity, would avoid such a problem. Nevertheless, the claims that Davidson makes along the lines that 'there is no manual for determining what a metaphor 'means' or 'says'' (and that there couldn't be one) are quite compelling. We are confronted again with our familiar dilemma. But while Searle was somewhat indecisive in choosing which horn to accept, before going for systematicity, Davidson makes an unequivocal choice of the opposite. Not surprisingly, the above example is an apparent embarrassment to his position arising from the option he chose to oppose. How is it then, that these two philosophers manage both to make such convincing claims, yet these claims are conflicting?

2.2.3 Settling the dilemma

The answer lies in their way of presenting the problem from two different perspectives, each giving credibility to one view rather than the other. The way Davidson presents it, which I will call a forward looking approach, is the following: 'given a metaphor, is there any particular proposition that can be systematically associated with it?'. To this, I agree that the answer is 'no'. The way Searle had presented the problem when arguing for the systematicity that metaphors display, or for that matter the way that I have presented the exam-

[6]Max Black discusses in extensive detail this example in his paper 'How Metaphors Work: A Reply to Donald Davidson' (Black 1978). The point I'm making about it, however, is worth being contrasted with Black's. He was arguing that there must be some content associated with this metaphor. I am only asking for criteria to distinguish what content *might* be associated with it, from what clearly isn't associated with it.

ple above, states the problem in what I shall call a backward looking approach. 'Given a metaphor and a particular interpretation, is there a relation between them that systematically applies to other metaphors and their given interpretation?'. I tend to agree with Searle that, at least in the case of establishable metaphors, their established interpretation (established for a community, or for an individual, as Davidson's approach would demand), is systematically related to their literal meaning.

These two perspectives are not contradictory. Let us assume, as I've suggested in the first chapter, that a metaphor has a *range* of propositional effects (or interpretations) associated with it. This range is restricted by some constraints, but it is not given which of these effects is the particular effect that a metaphor is going to have. A proper metaphorical effect can be defined as one that accords with the relevant constraints. A forward looking approach of such a situation will tell us that, for a given metaphor, there is no particular effect it is going to have. A backwards looking approach can still consistently maintain that given a metaphor and a propositional effect, we can determine if this effect is within the given constraints, i.e. if it is a proper metaphorical effect. We will be able to point out, in such a case, to a relation that holds between the metaphor and its effect which holds between other metaphors and their effects. That, without committing ourselves to the view that for any given metaphor there is one particular effect that it must bring about.

Given this description, we see that there is no longer a problem with assigning to these effects a propositional content, something that Davidson was reluctant to do. At least in the case of establishable metaphors, that do acquire a propositional content as they become established, we must assume that this content was one of the effects (possibly among many other effects, some of which may be non-propositional) that the metaphor had. The metaphor, presumably, acquires a content when it becomes established by one of its effects becoming

uniquely associated with it at the expense of the other effects that it originally had[7]. So, the propositional content of an established metaphor must have been one of its effects to begin with.

2.2.4 Re-evaluating Davidson's argument

Does the suggestion above answer Davidson's worries about giving a semantic account of metaphor? It should be noted that he does not dispute the idea that some of the metaphorical effects have propositional content in the first place. His actual arguments are stated in the two following quotes:

> 'The central error about metaphor ... is the thesis that associated with a metaphor is a *definite* cognitive content... This theory is false as a *full* account of metaphor, whether or not we call the purported cognitive content a meaning'.
>
> '...in fact there is no limit to what a metaphor calls to our attention, and much of what we are caused to notice is not propositional in character'
>
> (pp. 262–3, my italics)

What Davidson goes against, we see, are two things: first, the idea that there is a definite propositional content that can be associated with a metaphor. Second, he claims that an account of metaphor that mentions only the propositional or cognitive aspects of it, is incomplete. In other words, he is saying that a full account of metaphor is forced to discuss non-cognitive effects as well as the cognitive ones. This means that such a full account must extend outside the domain of semantics, that is supposed to, and equipped, only to deal with things that have a propositional content[8].

[7]Nelson Goodman makes a similar point in his paper 'Metaphor as Moonlighting' (Goodman 1978): '... if when "burned up" becomes a literal term for angry people, it has the same application as when metaphorical, then its metaphorical application must have been different from its other (original) literal application to things consumed by flame' (*ibid*, p. 177).

[8]Some philosophers (e.g. Hesse 1987, or Riquere 1978) would dispute the supposition that

I shall begin by evaluating Davidson's latter claim. Notice that this claim does not show that at least a partial account of metaphor cannot be given in terms of propositional content. The claim that *this* cannot be done was his first claim, on the indefiniteness of content, and we shall discuss it shortly. Now, a semantic account of metaphor, as partial as it may be, need not account for those non-cognitive aspects of metaphor anymore than a semantic account of literal meaning should account for the non-cognitive effects that surely every utterance has.

By this argument, then, Davidson hasn't shown that a semantic account, given in terms of propositional content, is impossible. But he adds another dimension to this claim. Supposing that we *could* assign a new meaning to a metaphor, do we want to do this? He argues that if we assign a new meaning to a metaphorical utterance, we get a theory in which 'to make a metaphor is to murder it' (p. 249). A live metaphor is distinguishable from a dead one, according to Davidson, exactly by the non-cognitive effects—images, visions, etc. —that it has. But if a metaphor just means something else, no such effects are to be expected, as in fact is the case with dead metaphors that loose their metaphoricality and are just a case of plain ambiguity[9]. Assigning a new meaning to a metaphorical utterance, however, is killing the metaphor only if you assume that the meaning of a metaphor must include everything, cognitive and non-cognitive effects together. This goes against Davidson's own conception of the limitations of a semantic theory. If all one is saying is that a metaphorical utterance has a new propositional content, by this one is not committed to any view whatsoever as to the non-cognitive effects that go or do not go along with this new meaning. One could coherently claim, for example, that the semantic fact that an expression just got a new meaning has non-semantic effects different from cases where this meaning had been pre-established. This claim, however,

cognitive content must be stated in propositional terms. I will avoid entering this dispute, and I will try to answer Davidson in his own terms, granting this supposition.

[9]Cooper, it should be noted, disputes this alleged fact. See (Cooper 1986), Chapter 3.

would not be part of his semantic theory. Davidson himself, in fact, acknowledges such a possibility. In 'A Nice Derangement' he treats puns on par with his treatment of malapropisms, i.e. by assigning to them a new meaning in the passing theory. In the paper on metaphor, he explicitly says of an example that 'This is a legitimate device, a pun, but it is not the same device as metaphor' (Davidson 1986, p. 250). Nevertheless, it cannot be denied that even though a pun does it work by *meaning* different things at once, it surely 'has a point' and has effects that are additional to its semantic aspects.

I believe it is fair to conclude that the considerations that have to do with the non-cognitive effects of metaphor do not have bearing on the plausibility of offering a semantic theory of meaning change in metaphor. Davidson's remarks, about the non-cognitive aspects of metaphor, have significance only in reminding us that metaphors have other, important aspects, that cannot be accounted for by a semantic theory of metaphor. It is somewhat ironical that Davidson needs to be reminded that a semantic theory can be stated independently of a full-fledged theory of understanding, where understanding includes all the nuances and implications of communication.

Let us look now at Davidson's first claim, that no definite propositional content can be associated with a metaphor. The claim can be sharpened by looking at the following quote:

> 'If what the metaphor makes us notice were finite and propositional in nature, this would not in itself make trouble: we would simply project the content the metaphor brought to mind on to the metaphor'.
>
> (pp. 262–3)

Davidson anticipates the move of assigning to a metaphor a multitude of propositions, but as we've seen above, says that this won't work because the

metaphorical effects, even if propositional, are not limited. This does seem to block claims about metaphors' having a particular content. A possible defense of such a position might revert to explanations in terms of vagueness of application, i.e. to saying that there is a limited, though not sharply defined, class of propositions associated with each metaphor that might be considered as their meaning[10]. Vagueness is a wide-spread phenomenon covering many cases of literal communication, and thus vagueness in itself is not a straightforward reason to avoid the assignment of meaning to an expression. I think, however, that no such steps to evade indeterminacy should be sought for. Indeterminacy should be accepted and highlighted, as it is the source of explanation for the aspect of semantic novelty that metaphors display. At the same time, as my proposal about metaphorical constraints suggests, indeterminacy doesn't imply chaos. It can be allowed without giving up the possibility of a systematic account that provides the necessary distinction between proper metaphorical effects, and other effects that a metaphor might have.

2.2.5 Meaning, or just content?

We have established that Davidson's arguments cannot pose an objection to an account of metaphor given in terms of propositional content. In that we have moved away from his claim that we cannot talk of metaphorical meaning because we cannot meet the minimal requirement of accounting for meaning in terms of propositions. Now we are left with the question of what other features a theory of meaning must have. Here we must meet two further objections that Davidson has raised against a notion of metaphorical meaning.

The first of these objections has to do with the explanatory value of an ad-hoc explanation. Davidson says:

[10]Sperber and Wilson, in their book 'Relevance' (1986) can be interpreted as arguing for such vagueness, though not as a response to this particular problem.

> 'It is no help in explaining how words work in metaphor to posit metaphorical or figurative meanings, or special kinds of poetic or metaphorical truth. These ideas don't explain metaphor, metaphor explains them. Once we understand a metaphor we can call what we grasp the 'metaphorical truth' and (up to a limit) say what the 'metaphorical meaning' is. But simply to lodge this meaning in the metaphor is like explaining why a pill puts you to sleep by saying it has a dormative power'.
>
> (p. 247)

First, a subsidiary remark: it seems as if Davidson is confounding an argument about an explanation's being ad-hoc, with an explanation's being circular. The pill analogy is about circularity, the claim concerning metaphorical meanings is about their being ad-hoc notions. Now, the first and most decisive objection to Davidson's argument is that he, after his paper 'A Nice Derangement', cannot afford to make it. In this paper he has argued that literal interpretation is itself many times ad-hoc. His position is that meaning is not what explains interpretation, but the other way around. He cannot, therefore, reject the notion of metaphorical meaning on the grounds that it is ad-hoc.

Davidson can be answered, however, in terms external to his own theory. His argument seems to go against my suggestion about the backward looking approach. The suggestion was that which systematic relation holds between a metaphor and its established meaning can be identified only for a *given* metaphorical meaning, without being able to predict the meaning of a given metaphor. But saying that this is an ad-hoc explanation doesn't mean that it is not an explanation. Explanations of evolutionary processes, for example, are just of this kind. They can explain, in an illuminating way, how certain forms have developed, through selection processes, from earlier forms. Nevertheless, such explanations do not have, or at least have a very restricted, predictive power (this analogy will be developed in detail in Chapter 7). Our explanation here of metaphorical meaning, then, isn't vacuous.

I expect, however, that Davidson would still reject this sort of explanation given as a proper part of a theory of meaning. Under his conception, such a theory does have to determine, for each sentence, a *particular* content[11]. When he says that 'there is no manual for determining what a metaphor 'means' or 'says'' (p. 245) he is contrasting it with literal meaning, for which a theory of meaning can be viewed as just such a manual.

Why is Davidson insisting on finding a theory of meaning for a language that strictly determines the meaning of each and every utterance used in successful communication? It is because he assumes that such a theory must explain the notion of linguistic competence—the actual processes that underlie successful communication. He claims to have falsified 'the assumption that communication by speech requires that speaker and interpreter have learned or somehow acquired a common method or theory of interpretation—as being able to operate on the basis of shared conventions, rules, or regularities' (Davidson 1986, p. 446). With this claim I agree, and I have endorsed it in my criticism of Searle's description of the "psychological" processes that underlie the ability to understand metaphors.

But this doesn't put an end to the need to explain what language is, and what it is for words and sentences to have their own meanings. Presumably, the connection between a notion of language, in the sense that English, French or Hebrew are languages, and actual communication processes is much more complex than just the idea that knowing one of the languages is a sufficient, or even a necessary condition for communication with people who are said to speak this language. As Dummett (from whom the latter point is derived) puts it: 'whatever force his [Davidson's] arguments may have, they cannot sustain the bald conclusion [that there is no such thing as language], but cry out for

[11] In cases of ambiguity, it can determine a limited range of actual contents. This is different, however, from the explanation we are discussing, where the metaphor acquires (in becoming established) one particular, actual content from the indefinite range of possible ones.

some account of an indispensable concept' (Dummett 1986, pp. 445–6). In the framework of such a theory of language and meaning, systematicities that are less than determining laws for every particular case are expected to be found, and can be a proper part of the methodology. In Chapter 4 I will present one such theory, and in Chapter 7 I will show how its application manages to allow a certain degree of systematicity in the characterization of metaphorical meaning. This characterization is a long way from specifying or determining exactly what this meaning is for each particular metaphor, in a way that fully explains metaphorical communication. I hope to show that such theorizing nevertheless makes sense and is valuable.

I have answered so far three of Davidson's objections for giving a semantic account of metaphor: (i) I have argued why metaphor is amenable to treatment in terms of propositional content. (ii) I have claimed why non-determinacy of content isn't a good enough reason for rejecting a semantic account. (iii) I have shown why an ad-hoc attribution of meaning does not exclude the plausibility of offering a semantic account of metaphor. The final consideration for why metaphor is not a semantic phenomenon that I will try to meet is derived from 'A Nice Derangement'. Davidson argued there, through the example of Jones' referential use, that even when a definite cognitive content can be identified, it may still be a case of higher order intentions, rather than first order, meaning-determining intentions. Similarly, he may argue that even if a determinate cognitive content can be associated with a metaphorical use, still the first order intentions may remain unchanged.

I must say that I cannot see what are the principled reasons for refraining from attributing the intended message—both in Jones' case, and in the case of metaphorical use—to the first intention of the speaker. If Davidson had argued that in such uses the hearer wouldn't want to make changes to his long-term, prior theory, I could agree. A hearer can recognize in such cases that the

speaker was aware that he is using his words in an awkward way, and should not be expected to continue to use them in that way on further occasions. For example, Jones' hearer wouldn't think, after hearing him say that Smith's murderer is insane, that from now on Jones is going to use 'Smith's murderer' as the name of the person which he had in mind. Davidson considers, however, even changes to the passing theory as semantic changes. But, since the passing theory is a disposable, one-off theory for interpretation, what stops the hearer from making this semantic change? Why should he not decide that on that particular case, 'Smith's murderer' should be interpreted as referring to Mr. Ripper? Similarly, why should a hearer of a metaphor refrain from making a one-off change to his passing theory for the speaker? I think, therefore, that a claim along this line again falls short of establishing that metaphors should not be treated as a semantic phenomenon, according to Davidson's own (rather peculiar) notion of semantic change.

Let us now turn to have a look at another theory of metaphor, suggested by M. Hesse. Her theory is far removed from talking about intentions of individuals and processes of interpretation, and it will not satisfy Davidson's demand for a full theory of communication. But, since Davidson himself has contributed so much to our realizing that such a theory of communication is pointless as a theory about the notion of language, we can be open minded and see what other forms of theorizing can throw light on this notion, as well as the notion of metaphor.

Chapter 3

Hesse on Metaphor

Hesse's account of metaphor[1] is based on the 'interaction view of metaphor', that was presented by Max Black in (Black 1955). Her following claim, and two of its implications, will be the main concern of my discussion of her view. She says that 'a metaphoric expression used for the first time ...is intended to be understood' (p. 164), and is capable of being misunderstood. The two implications of this claim are: (i) that metaphors must have a cognitive content, and (ii) that there must exist rules for metaphoric use. Hesse's position in general, and in particular these claims, are of interest to us as they represent a choice opposite to that which Davidson has taken in the debate about regularity versus irregularity of metaphorical meaning. As we shall see, Hesse's approach to the notion of semantics, drawn from her idea of what a theory is in general, and in particular her view of what a semantic theory should look like, are very different from Davidson's. Accordingly, one of our main tasks will be to clarify what really is at stake in the debate between Davidson and Hesse, over and above the dogmatic claim that a metaphor does, or doesn't have a regular meaning. I shall begin by presenting Hesse's theory of metaphor, and then discuss it,

[1] I will be presenting mainly her views as were expressed in 'The Explanatory Role of Metaphor' (Hesse 1966). All page references in this chapter, unless otherwise stated, are from this paper.

concentrating on the above claims.

3.1 Exposition

Hesse's aim in her discussion is to show that metaphors have an indispensable role as part of scientific explanations, one that cannot be replaced by use of a purely literal language. To do this, she begins by presenting a theory that explains what metaphors convey. She then shows how this view of metaphor is able to solve problems that the classical deductive model of scientific explanation has encountered.

A metaphoric use of language, Hesse claims, consists of describing one system, labelled the 'primary system', with words normally used in connection with another, 'secondary system'. A 'system' denotes the referents of descriptive statements. Each system has 'a set of associated ideas and beliefs that come to mind when the system is referred to. These are not private to individual language users, but are largely common to a given language community' (p. 159). The notion of 'meaning' that Hesse uses in her discussion is to be understood as 'an inclusive term for reference, use, and the relevant set of associated ideas' (p. 160). Accordingly, claims about shifts in the meaning of an expression can be shifts in any of the above.

For an expression to be a metaphor, it must be patently false or absurd if taken literally. However, 'There is initially some principle of assimilation between primary and secondary systems' (p. 160). By this, Hesse seems to say that there is some relation of analogy, or similarity, between the two systems. Hesse warns against two misinterpretations of this claim. First, she says, it is not the case that any two systems can be combined in a metaphorical statement. That is, not every absurd combination will result in a metaphor that is explana-

tory, in the sense of making a claim with a content that is refutable. She admits that in poetry this might not be the case, but distinguishes poetical metaphors from scientific ones. The latter, in being explanatory, must have refutable content, and thus impose restrictions on the viable metaphorical combinations of system descriptions. Hesse, in other words, is taking a realistic stance towards the notion of similarity. There is a fact of the matter of whether two systems are similar, rather than just a subjective act of imposing a decision to find such similarities. Also, claims about a similarity that holds between two systems are not vacuously true. Even if it is the case that any two systems can be said to be similar in some respect, this does not license just any arbitrary categorization to be made, as some systems are more similar than others, and they are more similar in some respects than in other respects. Second, Hesse points out that the demand for a relation of similarity of some sort to exist between the systems to make a metaphor meaningful does not imply a 'comparison view' of metaphor. That is, it is not the case that there must exist a literal statement that can fully replace the metaphor by specifying the similarities between the two systems. One reason for making this claim is that 'as long as a model is under active consideration as an ingredient in an explanation, we do not know how far the comparison extends—it is precisely in its extension that the fruitfulness of the model may lie' (p. 162)[2]. In other words, the way in which two systems are similar is a matter to be revealed by further investigation, and is not just given by the making of a metaphorical statement.

'The metaphor works by transferring the associated ideas and implications of the secondary to the primary system. These select, emphasize, or suppress features of the primary ...' (p. 163). This is one sense in which the metaphor causes a change in meaning. The meaning of an expression, we have seen, has to do with the set of ideas associated with it. The shifting of ideas from one

[2]Hesse identifies models and scientific metaphors in this paper, so this claim can be taken as one about metaphors, just as about models

domain to the other thus causes meaning change in the words used in describing both systems. With the primary system this is obvious, as it is being described in terms that are not normally associated with it. But even the descriptions of the secondary system change their meaning, by the very fact that they have been associated through their use in the metaphor, with the primary system. This claim about meaning change is another reason why the interaction view is incompatible with the comparison view, mentioned above. The use of a metaphor results in a change of meaning; the use of a literal statement of similarities does not have this effect. A literal statement, thus, must be describing, in using the words it does without a change in their meaning, something different than what the metaphor describes.

Even though a metaphor changes the meaning of the words it consists of, this change is a definite one: the metaphor, already on its first use, is intended to be understood. It can also be misunderstood, if conveying something other than that which was intended. Metaphor, therefore, has a cognitive content, rather than just being 'a wholly noncognitive, subjective, emotive, or stylistic use of language' (p. 164). A metaphor, like a model or a theory, has an objective content. It can be rejected by showing some of its implications to be empirically false, even if these are implications that the originator of the metaphor did not think of. Accordingly, there are objective rules of metaphorical use that determine what a metaphor means. In this sense the distinction between literal and metaphorical interpretation cannot be that the first is rule governed while the latter isn't: both are rule governed, by rules that are different but interdependent (since literal meaning is changed by metaphorical uses). Also, language is dynamic in the sense that metaphorical expressions may become literal, and vice versa. Over all, the interaction view does not accept a clear distinction between literal and metaphorical expressions, but rather one that is relative to context.

The interaction view described above raises the question of what is the ref-

erent of the metaphorical expression. Prima facie it seems that the primary system is the metaphor's referent, but then how could it be claimed that the patently false and absurd statement made about it in a metaphor is in fact true of this primary system? More precisely, the question is how *could* these descriptions be true, as it is not always the case that metaphorical claims are in fact true (just as not every literal statement is always true). The answer is that metaphorical statements are false and absurd only if taken literally. But, since the interaction view allows for the metaphorical meaning to be different from the literal one, the metaphorical statement can be true of the primary system it refers to. By this Hesse seems to be saying that the referent of the metaphor, while still being considered the primary system, has somewhat changed. This is another claim about meaning change, on top of the one mentioned earlier, that involved change only of the ideas associated with the two systems. According to the present claim, metaphor initiates also a change in reference. Given that such a change is assumed, Hesse says that 'there remains the question what it is to identify the referent of the metaphoric expression or model with the primary system' (p. 168). In other words, there remains the question of explaining in what sense the new and somewhat different metaphoric referent is still related to the original referent.

To answer this question, Hesse points out again the difference between poetic and scientific metaphors. In the first it may be essential that the metaphorical statement remains contradictory and false, for the metaphor to achieve its effects which are not of the cognitive sort. A poetic metaphor is not intended as an explanation, and thus there is no need to achieve consistency between its parts. For example, there is no need to reconcile or choose between two conflicting implications that a poetic metaphor may have. Scientific metaphors are different: 'they are meant to be internally tightly knit by logical and causal interrelations; and if two models are found to be mutually inconsistent, this is not taken ... to enhance their effectiveness but rather as a challenge to reconcile

them by mutual modification or to refute one of them' (p. 169). If I understand correctly, the relevance of the distinction between poetic and scientific metaphors to the question of what it is to identify the reference of the metaphoric with the primary system is in showing that meaning variance should be an *acceptable* answer as to how metaphors function in an explanation. That is, it is not directly answering the question of how the new referent is associated with the old one, but rather argues that it makes sense to talk of a meaning change, because in scientific metaphor contradiction is not an essential feature of the metaphor. Quite the contrary: scientific metaphors are to avoid contradiction, which justifies the need for a change in meaning.

I will postpone the presentation of Hesse's positive arguments for supporting her view that scientific metaphors are best explained by the meaning change that the interaction view of metaphor describes. Those basically involve showing how flaws in the deductive model of scientific explanation are overcome by supplementing this model with the interaction view of metaphor. Her conclusion, however, is that while under the deductive model, changes in vocabulary cannot be rationally accounted for, the metaphorical view of explanation does show these changes to be rational. That is because 'rationality consists just in the continuous adaptation of our language to our continually expanding world, and metaphor is one of the chief means by which this is accomplished' (pp. 176–7).

3.2 Discussion

It is somewhat problematic to compare Hesse's account of metaphorical meaning with the former two accounts that we've encountered, one particular problem being her wide conception of what meaning is. For Hesse, we have seen, meaning is an inclusive term for reference, use and the associated ideas that a description

of a system carries. For Searle and Davidson, meaning is explicitly *contrasted* with use, and even more so with further notions such as associated ideas. A more specific complication that arises from this difference is that Searle and Davidson see the problem of metaphor very much as concerning the distinction between what words mean, and what particular speakers mean on specific occasions of communication. That is, the intentions of individuals seem to them very relevant to questions of meaning in general, and much more so in cases of deviant uses of language. For Hesse, on the other hand—maybe because she doesn't think that metaphors are deviant uses—the intentions of particular speakers do not seem relevant to the discussion of metaphor. Even when talking about the associated ideas and beliefs that the primary and secondary systems carry, she says they are 'not private to individual language users but are largely common to a given language community' (p. 159). She is in fact insensitive to the distinction between those subjective intentions, and the objective features associated with a meaningful expression. For example, when she says that metaphors are intended to be understood, and can be misunderstood, she describes a case of misunderstanding as one where a hearer failes to get at the particular speaker's intended meaning. This doesn't stop her from saying on the next paragraph that a metaphor can be rejected as false by attending to implications that the particular speaker didn't even think of, which depend on the metaphor's being inter-subjective rather than 'a private language of the individual theorist' (cf. pp. 164–5).

I think it is fair, however, to take Hesse as wholeheartedly committed to an objective view of metaphorical interpretation, where a speaker can only use metaphors in a way that is prescribed by the metaphorical rules of interpretation, his intentions being irrelevant to what the metaphor itself means. In that respect Hesse is clearly offering a view opposite to Davidson's, and different from Searle's (Searle is not, at least intentionally, committed to a view of metaphors' having an objective meaning, in the same manner as words have an objective literal

meaning). But even if this point is relatively clear, we shall see that further differences between Hesse's and Davidson's general approach to language may cause much confusion in clearly stating the discrepancies between their accounts of metaphor.

I have said that the basic problem with accounts of metaphorical meaning is the tension between claims about the regularity that holds between metaphorical expressions and their meaning, and the opposite intuitions that metaphors are irregular in this respect, being novel and unprecedented uses of language. In the discussion of Davidson's theory of metaphor, we have seen that his taking sides on this issue inevitably raises objections from the other pole of the above mentioned contrast. Accordingly, we may expect that Hesse's choice of the 'metaphors are regular' option should elicit objections from the 'irregularity' camp. What are the objections to her position, following this consideration?

3.2.1 The metaphoric version of the deductive model of scientific explanation.

Hesse's thesis in her paper 'The Explanatory Role of Metaphor' was that 'the deductive model of scientific explanation should be modified and supplemented by a view of theoretical explanation as metaphoric redescription of the domain of the explanandum' (p. 157). Her arguments in favor of the metaphoric view of scientific explanations are intended to show how adapting this view solves problems associated with the classical deductive account (e.g. how a body of theory can be rationally expanded and how the formal, mathematical apparatus of the theory is connected to its empirical claims). This success in application should serve also as an indirect support for the suggested account of metaphorical meaning. I will argue, however, that Hesse's claim that there are rules for metaphoric interpretation as presented in her paper, may be so interpreted that it fails to 'modify' or 'supplement' the classical deductive view. Instead, it may

be seen just as an addition of more machinery of basically the same kind to the existing theory, that doesn't really enhance the plausibility of the original deductive model.

Hesse seems to draw the following picture of metaphoric interpretation: alongside with the rules of literal interpretation, there is an additional set of rules, those of metaphoric interpretation. Accordingly, she says,

> 'It is not possible to make a distinction between literal and metaphoric descriptions merely by asserting that literal use consists in following linguistic rules. Intelligible metaphor also implies the existence of rules of metaphoric use'.
>
> (p. 165)

The two sets of rules are said to be interdependent, in what I understand to be the following way: the metaphoric rules dictate, for a given metaphor, the way by which the meaning of the terms involved should vary in their meaning. In that they modify the rules of literal interpretation. This is what enables Hesse to claim that already on the first instance when a metaphor is used, its meaning is determinate. The metaphor has, thus, an objective, specific, cognitive content which is falsifiable in the same manner as any literal statement is subject to falsification. We may compare this claim to Davidson's description of the communication process involved in understanding malapropisms: a metaphor, like a malapropism, acquires a new, non-standard meaning on the instance of its being uttered. The clash between their view's, then, is over the question of whether metaphors do have such a specific content, and whether the transition from standard meaning to the new meaning is rule governed or not (Davidson, recall, argues that while malapropisms do have a specific content, there are no rules that determine how this content can be recovered from what had been uttered).

Now, let us see what work this view of metaphor is supposed to do for the

deductive model of scientific explanation. The first problem with the deductive model is that in many cases the statements deducible from the scientific theory (the explanans) have only approximate fit with the statements of observation made by an empirical study of the domain of what is to be explained (the explanandum). In a deductive theory of explanation committed to meaning invariance, this should be enough to falsify the given theory. But since such a description is not adequate of the actual scientific practice, this renders the deductive model false.

Hesse claims, however, that the deductive model can be salvaged if it allows for a metaphoric shift to take place in the meaning of the terms used in the description of the explanandum, so that they do make true the statements deduced from the explanans. This may seem a dangerous move, since it may imply that any theory can be sustained by making the appropriate shifts in the meaning of the descriptive terms. But this is not the case, as what metaphoric shifts can take place are strictly determined by the rules of the metaphoric interpretation.

Has anything substantial been added to the deductive model by this manoeuvre? Not if our above interpretation of Hesse's theory has been correct. All that the deductive model was missing, according to Hesse's suggestion, was an additional *logical* ingredient that would enable us to logically deduce from the explanans a correct description of the explanandum. Admittedly, this description will involve predicates with a new, or extended meaning. Nevertheless, what these new elements will denote, and thus what the theory is saying, is strictly deducible from the explanans together with the rules for metaphoric interpretation (or better, maybe, rules for metaphoric re-interpretation). Now, the rules of literal interpretation are required anyway to evaluate the validity of any literal descriptive statement, and Hesse herself implied that there is no important distinction between the literal and the metaphoric rules of interpretation. It seems therefore that there is no important distinction to be made between

the classical and the 'modified' accounts of deductive explanations in science. The modified account just says that the problem with the classical model had been an incomplete account of language. Once this deficiency is corrected by adding the missing rules of metaphoric interpretation, the deductive model is made good. That is, the fit that *seemed* to be only approximate, was in fact a perfect fit, given the correct interpretation of the theoretical claim.

Consider now Hesse's solution to the second problem with the deductive model. The problem concerns the matching of corresponding terms used in the observation language with those of the separate, theoretical language. If these are taken to be two separate languages, further rules which are not deducible from the explanans alone are required for assessing the validity of statements made in the theoretical language against statements made in the observation language. Hesse urges instead that there is only one language, that of observation. Accordingly, the validity of any statement is examined directly against reality, where the combined literal and metaphoric rules of interpretation determine what exactly the statement is saying of that reality. Again, this view is not saying much about the deductive model. All it is saying is that if you get your linguistic rules right, the deductive model works. Metaphor doesn't really supplement the deductive model, it only supplements the linguistic theory that accompanies this, or any model of explanation.

It may be argued that there is nothing wrong with the above account of the deductive model. Hesse may be taken to argue that the deductive model is, in fact, correct, and all that was needed as a supplement was in fact just the introduction of these further rules of metaphoric interpretation. I shall thus sharpen the above argument by attending specifically to the account of metaphor assumed by it. The problem with this account is that under it the distinction is lost between a new element of a language being introduced and its meaning defined, versus its being used, once it is so defined, to make a statement about

some facts. Hesse does not deny that a metaphorical statement involves new elements of language. In fact she explicitly argues that this is precisely what metaphors are. 'In the metaphoric view', she says, 'it is to be expected that the original observation language will both be shifted in meaning and *extended in vocabulary*' (p. 176, my emphasis).

Is metaphor, then, a sentence that is introducing a new vocabulary, or is it one that is using a new vocabulary? If metaphors do introduce a new vocabulary, and this is done by the new meaning being a function of the original, literal meaning of the words used, a metaphor can be compared with a definition. Definitions too are cases where new elements are introduced to a language by using an existing vocabulary. But a definition is not making any empirical claim, and cannot be evaluated as true or false, whereas metaphors are making empirical claims, according to Hesse. A metaphor must be, therefore, a case of using a new vocabulary for making a factual statement. But then, where did this new vocabulary get its meaning from? If it just gets its meaning from the combination of existing words in a new way, according to a prescribed mode of combination, a metaphor has a meaning which is "new" only in the sense that a literal sentence that happened not to be ever uttered before is "new". According to such a view, there would not be an essential difference between a new theory stated in literal terms, and one stated in metaphorical terms. But Hesse's modification of the deductive model was intended to show that there was such a difference.

It may again be argued that presenting the problem in this way, such that it assumes a distinction between meaning and use, is a distortion of Hesse's view: she does not accept this distinction to begin with. According to Hesse, the very fact that the metaphor was used in a new way, defines the new meaning of its terms. That is, metaphors are a case where the new terms are introduced by being used. While it is true that Hesse doesn't accept this distinction, that

doesn't really solve the problem of how the metaphor gets its meaning. Can a metaphor (or its elements) be said to have a definite meaning just by being used *once*? And, if its meaning is not defined before it is being used, how is it determined what the metaphor is being used for? Searle, we have seen, tried to answer this latter question by letting the use be determined by the speaker's intentions. This, however, turned out to be a problematic solution, and anyway it is not the one Hesse opts for. But even if we assume that there is some objective way to determine what the metaphor is being used for, there still remains the following question. If it is the use that determines the metaphor's meaning, what role do the rules for metaphoric interpretation have? Is it those rules that determine its meaning, or is it the use? The claim that use determines the meaning makes the metaphoric interpretation rules redundant, so such an account is inconsistent with the earlier claim that it is the interpretation rules that determine a metaphor's meaning.

All this debate, of course, is again the emergence of our basic dilemma. It shows that even if the distinction between meaning and use is dropped, still it is not clear how a metaphor can be attributed with a specific meaning on the first occasion of its being uttered, without effectively denying that metaphors are new and irregular uses of language. This, we have seen, is not a conclusion at which Hesse would want to arrive. If the above interpretation of Hesse's account was correct, therefore, these seem to be conclusive reasons for rejecting it. I think, however, that Hesse was in fact arguing for a different view of metaphor, one that answers these difficulties.

3.2.2 An alternative interpretation of Hesse's view on metaphor.

Our problem with the above interpretation of Hesse's view was with the claim that on its first appearance, a metaphor has a definite cognitive content that

is determined by the rules of metaphoric interpretation. But now let us look again at the passage that was quoted above, where she says that the comparison view is wrong because 'as long as a model [or a metaphor] is under active consideration ...we do not know how far the comparison extends—it is precisely in its extension that the fruitfulness of the model may lie' (p. 162). This could certainly be interpreted as saying that for a while, it is not determined what exactly the metaphor means. In other words, the meaning of the metaphor (or its elements) is open-ended, rather than determined, during that stage. In fact, the above must be a claim about the model or metaphor's *meaning*, not just its 'fruitfulness'. Only as such can it serve as an argument against a similarity statement's being synonymous (i.e. equal in *meaning*) with the model.

This alternative picture fits nicely with Hesse's inclusive view of meaning as reference, use and associated ideas. For, apparently, what use the metaphor serves exactly, and what ideas become associated with a newly referred to system must take time, and repeated applications in different domains of thought (especially if these are to be shared by a linguistic *community*). It seems, then, that Hesse is not quite committed to the view that metaphors do have a definite meaning on their first appearance. But this conclusion appears to be problematic, because she did explicitly say that metaphors are intended to be understood, and thus have a cognitive content, on their first appearance. Moreover, if it is this 'active consideration' process that stabilizes the metaphors' meaning, the question arises again as to what role do the rules of metaphoric interpretation have? Is it the rules, or is it the active usage that determine the meaning of a metaphor? Before addressing these problems, let us look further at the nature of the rules of interpretation for metaphorical statements that Hesse argues for.

In a later paper Hesse says the following:

> 'It may well turn out that nothing but broad generalities are possible in such an account [of theory change], supplemented by detailed study of particular cases'.
>
> (Hesse 1987, p. 305)

And again:

> 'The theory of metaphoric meaning I have projected would provide such predictability for metaphoric and literal talk. In the case of metaphor it will not be tight and univocal predictability, because the nature of metaphor is to be allusive, evocative, and partly ambiguous. But ambiguity is not total—we *do* know how to respond to tropical talk'.
>
> (*ibid*, p. 309)

What emerges is that there is an interesting difference between literal and metaphorical talk, even if this difference is, as Hesse insists, just a matter of degree. While in the case of literal talk the meaning of an expression can be uniquely identified, metaphorical talk sets only broad limitations on the possible interpretations of an expression. Such a view is very much like the one I've been suggesting in the earlier chapters. That is, while a metaphor doesn't have a determinate meaning on its first appearance, nevertheless the literal meaning of its elements place broad constraints on the possible meanings it may acquire as it is later being used and becomes established.

This alternative understanding of Hesse's view on metaphor now makes sense of her claims that she was amending the classical deductive model. It is no longer the case that the meaning of a metaphoric expression is definite in a way that makes purely deductive explanations plainly right. Instead, the metaphoric interpretations open the possibility for a restricted range of possible meanings, that may eventually make what initially was only an approximate fit into an exact fit of an empirical claim. In other words, the metaphoric rules of interpretation at the same time open and delimit the change of meaning in the

terms being used, in a way that allows for first approximations to have a role in scientific explanations, but without implying chaos.

Our earlier problem, of whether it is the use or the rules that determine the meaning of a metaphor is now readily answered: it is the combination of both. The rules set restrictions on the interpretation, leaving room for ambiguity, but one that is not total. The use, over time, fixes the established meaning to one possibility out of the open-ended range. What about Hesse's claim that metaphors do have a content on their first appearance? I think that making this claim, together with the unexplicated use of 'metaphoric rules' alongside with 'literal rules' are an unfortunate source of confusion. However, I believe we would be correct to interpret this claim again as saying that metaphors, on their first appearance, have a cognitive content only in the sense of having a limited range of interpretations. In that way they are not meaning*less*, and can be misunderstood if interpreted in a way that is outside the range permitted by the metaphoric rules of interpretation. Another way of misinterpretation, that makes sense of Hesse's reference to the speaker's intended meaning, is that the ambiguity (or rather indeterminacy) allowed for in the case of a new metaphor may enable a speaker to intend one permissible meaning, and be interpreted as intending something else within that range. Such an account of misinterpretation allows reference to be made to the speaker's meaning while avoiding the pitfall that Searle has stumbled over. The speaker's meaning is relevant to which of the possible meanings the metaphor was used to convey, but the speaker cannot just intend to convey anything whatsoever.

Finally, it is interesting to compare another claim that Hesse makes, to the effect that there is a difference between scientific and poetic metaphors, with what I've been suggesting. I have said that we should expect to find a difference between one-off metaphors and establishable ones. Hesse's claim that poetic metaphors are not constrained in their interpretations, and may be essentially

contradictory, can be taken as a claim about at least part of what makes a metaphor non-establishable. Conversely, her suggestion about principles of similarity that hold between the two systems of a scientific metaphor point out some characteristics of establishable ones.

3.2.3 Hesse vs. Davidson.

Now that we've gained a better understanding of Hesse's theory of metaphor, we are in a position to go back and compare what she actually says with Davidson's position, that was supposedly the complete opposite to her own. One pole of debate is between the view that semantics should be, as Hesse puts it, about 'a frozen stage of natural language, and all it can deal with is a discrete series of such stages' (Hesse 1987, p. 297), and the view 'that the problem of change is a proper part of its brief' (*ibid*, p. 298). It was our forcing of the static view of semantics on Hesse's account that had first mislead us in our interpretation of her account. In our discussion later it emerged that Hesse was in fact not concerned with such a static account of language. Her claims about metaphor resulted from a dynamic view of semantics that aims at describing not only what an expression means at a certain time, but also how meanings change, and how language expands.

The second, and closely related issue, on which Hesse and Davidson disagree is whether a semantic theory must be such that it gives the exact meaning of every sentence in a language, or if instead a more general theory that may include broad restrictions on linguistic meaning can also be a proper part of semantics. The disagreement on the first question is, to some extent, a consequence of the disagreement on the second—I imagine that if Davidson had thought that meaning change could be given in logically rigorous terms, he wouldn't have denied that such a theory of meaning change has a place within semantics.

Similarly, he wouldn't have come to the conclusion that 'there is no such thing as language', because the shifting 'prior' and 'passing' theories together with the alleged rules controlling the shifts between them would come close enough to what he would consider a language. What he thinks is missing, of course, are the transition rules from one theory to the other.

It is interesting to see how, despite these disagreements, Hesse and Davidson are close in their observations: they are not quarreling over the facts, i.e. that language changes cannot be characterized by rigorously statable rules. It is just that Davidson thinks that for a theory to be a theory it must be so statable, and thus the theory of language change cannot qualify as a theory at all. Hesse, on the other hand, thinks that broader generalizations with only loose fit with each particular case can be treated as a theory, and therefore the theory of language change itself is one. Similarly with their view on metaphor: they do not disagree on the fact that a metaphor doesn't have a uniquely specifiable content on its first appearance, as literal statements may have, but rather an open-ended content. They only disagree on whether what can be said, more generally, about metaphorical content is worthy of being included in a theory of meaning.

There is, in fact, a third point of disagreement between Hesse and Davidson, which my way of presenting her views has suppressed. Hesse, in claiming that metaphors do have a meaning, is in fact arguing that seeing one thing as another, or organizing one system in terms of another, has a cognitive significance over and above just the set of propositions that are made true once the appropriate meaning changes in language take place. This, I think, is what Davidson has in mind when he says that a metaphor has 'a point' that it is making. But, Davidson says,

> 'Seeing as is not seeing that. Metaphor makes us see one thing as another by making some literal statement that inspires or prompts

> the insight. Since in most cases what the metaphor prompts or inspires is not entirely, or even at all, recognition of some fact, the attempt to give literal expression to the content of the metaphor is simply misguided'.
>
> (Davidson 1984, p. 263)

I agree with Davidson that there is an important aspect of metaphor that is not captured by just coupling a metaphor and a proposition that expresses some fact. At the same time, I agree with Hesse that whatever this insight may be, it certainly has a lot to do with cognition. Hesse is right in insisting that there must be more to cognition than just 'bare facts'. But I must admit that I fail to understand what this additional factor is, and how it is connected with the impoverished notion of meaning that deals only with facts and their expression in language.

The presentation of Hesse's view on metaphor in terms of my model of an indeterminate connection that holds between a metaphor and a range of meanings on its first appearance, and the narrowing down that occurs as the metaphor becomes established, has the disadvantage of glossing over that deeper point that she is arguing for. It has, however, the advantage of both making the discussion more intelligible, and of separating, within Hesse's account, those points that can be compatibly compared with the other accounts we've discussed. As we have seen, there are enough interesting questions to deal with within that simplified framework.

3.3 Conclusion

As expected, the attempt to understand Hesse's claim about the rules that govern metaphorical meaning within a static, synchronic view of language and meaning, has failed. Under this view, her claims that metaphors introduce

meaning change, or a new vocabulary, were inconsistent with her claims that metaphoric rules of interpretation determine a fresh metaphor's meaning. Coupled with our discussion of the deficiencies of Searle's and Davidson's accounts that tried to offer alternative answers as to the connection between metaphors and meanings, we may conclude that the synchronic view of language is faulty. We have seen, however, that interpreting Hesse's account as one that deals with language change, i.e. viewing it within a diachronic account of language, solved the difficulties presented by the first interpretation of her theory. Moreover, we have seen that a diachronic view of language made sense of both Searle's and Davidson's accounts. It preserved their basic insights, while solving the problems their accounts had with dealing with our basic dilemma concerning the meaning of metaphor.

In the second half of this thesis, I will present an alternative account of meaning and language, that highlights the claim that any facts about the meaning of linguistic expressions are intrinsically dependent on their history. Such an account, I will claim, provides us with adequate tools to deal with the problem of metaphorical meaning.

Part II

An Evolutionary Look at Meaning

Chapter 4

Millikan's Theory of Meaning

4.1 Introduction

In this, and the next chapter, I will mostly set aside the issue of metaphors and their meaning. This chapter will present a theory of meaning that was suggested by R.G. Millikan in her book 'Language, Thought, and Other Biological Categories' (1984)[1] and in the next I will try to defend her approach from her main critic, J. Fodor. The reason for having this rather detailed presentation of Millikan's theory is that in chapters 6 and 7 I intend to apply her approach to the question of metaphorical meaning. This approach, I will argue, manages to resolve the tension that the basic dilemma about metaphorical meaning has presented us with. The theory is not an easy one to grasp, frequently because of its esoteric use of familiar notions. Moreover, it uses those notions to make claims that are sometimes radical and counter-intuitive. In reading this theory, therefore, intuitions must often be set aside, and I shall do my best to explain Millikan's technical terms as I introduce the theory.

What's so special about Millikan's theory of meaning? Many things are, as

[1]Page references in this chapter are to this book, unless otherwise stated.

we shall see. But my reason for choosing it to clarify issues about metaphorical meaning have to do with particular strengths it has. One problem that the discussion of metaphor brings to the foreground, is the intricate relations between what linguistic expressions mean, and what their utterers mean by using them. Metaphor seems to walk on the very narrow and unstable border between the two, as was apparent in Searle's and Davidson's discussion of the phenomenon. The main problem with securing the line between the speaker's intentions and the expressions' meaning seems to me the fact that it is not made sufficiently clear what exactly makes words have their *own* meaning. Even when this is attempted, in Davidson's account, the meaning of expressions is characterized in a way that is very closely related and dependent upon the speaker's intentions. This leads, as we saw, to very radical conclusions such as 'there is no such thing as language', that can perhaps be avoided if an account can be given that will not tie linguistic meaning and speaker's intentions so tightly together. One advantage of Millikan's theory is in offering an account of the meaning of linguistic expressions that is, to a large extent, independent of any particular utterer's intentions.

Another issue that is brought to focus in a discussion of *fresh* metaphors (which according to Searle and Davidson are the only kind of genuine metaphors) is the relation between the meaning of linguistic expressions in the past, and their present meaning. The insistence on discussing only metaphors that are novel obviously assumes that there is much significance to the relation between the history of an expression and its present meaning, but unfortunately both Searle and Davidson hardly discuss this relation. This lacuna is closely related to the problem mentioned earlier, of not stating clearly enough what gives an expression its own meaning, for it seems that the history of an expression is a major determinant for what an expression means. This, again, is a strength of Millikan's account. As we shall see, her account of meaning places the full burden of explaining what an expression means on its history. Accordingly, the

question of what a fresh metaphor means will have to be answered in terms of the significance of the history of the linguistic elements used in a metaphorical expression to its present—and future—meaning.

Hesse's account, we have seen, fairs better on this latter consideration; it explicitly deals with meaning in terms of change over time. However, her account is not very clear on the first issue, of the relation between speaker's intentions and an expression's own meaning. Also, her characterization of meaning as reference, use and associated ideas is too vague. The notion of 'associated ideas', in particular, again blurs the distinction between an individual's grasp of meaning, and an expression's own meaning. Millikan's account helps to rectify these deficiencies in Hesse's account.

What is Millikan's motivation for giving an account of meaning that is independent of a particular speaker's intentions? Why not explain the meaning of sounds and marks on paper, as being derived from speaker's intentions? One important feature of meaningful objects that needs to be explained is their intentionality, their being *about* something. Millikan is interested in giving a naturalistic account of meaning in general, and of intentionality in particular, i.e. one that will explain these notions in non-intentional and non-semantic terms. In such an account explaining the meaning of linguistic objects in terms of speaker's intentions runs the risk of being circular, or inviting regress. How is the meaning of human intentions going to be accounted for in naturalistic, non-intentional terms? This consideration may support the motivation for explaining the meaning of linguistic entities independently of speakers' intentions. This alone, however, is not a good enough reason for avoiding a 'derived intentionality' kind of explanation. It is possible that a naturalistic explanation can be given for what it is for people to mean, or intend, without this explanation applying independently to linguistic entities. Those, in turn, can then be accounted for as having their meaning derived from that of the speaker. There

is another reason, however, why it is necessary to distinguish between speaker's meaning, and linguistic meaning. The reason is that speakers depend on words, as much as words depend on speakers. For a speaker to convey his ideas, there is a sense in which he needs to choose the *right* words for the job, and it is very difficult to explain this normative aspect of linguistic use if the norms are not set independently of the speaker. If, however, linguistic entities can be ascribed with meaning of their own, a speaker's use of them can be evaluated as coinciding with, or deviating from, their proper meaning.

So, Millikan wants an explanation of what it is for linguistic expressions to mean something, that isn't dependent on what any particular utterer means. How are we to find such an explanation? A possible strategy might be to try to find other objects, that are not produced or designed by human beings, that 'mean', or at least display characteristics that are as close as possible to those of meaningful linguistic objects. A preliminary question would then be, what are these characteristics that we're after? What is it for a linguistic expression to mean something? There are two major trends in the philosophy of language for trying to characterize what meaning is (as distinct from the question of what determines meaning). One discusses meaning in terms of correspondence between linguistic expressions and other entities in the world (e.g. Frege); the other discusses meaning in terms of what linguistic expressions are for, or what they are supposed to do (e.g. late Wittgenstein, Austin). The first approach can be seen as concentrating on the intentionality of linguistic expressions, on their being about other things; the second approach is focusing on the functional, or teleological aspect of linguistic expressions. As Millikan points out, both intentionality and teleology can be thought of as quasi-normative terms: an intentional entity is supposed to be about, or correspond in some way, to some other entity, but it doesn't always correspond as it should. A false sentence, for example, does not correspond to the way things are in the real world. A functional entity has a purpose, or a telos to fulfill. Again, it may fail to perform

its function, but nevertheless there is a sense in which it is supposed to perform it. For example, if A says to B 'sit down', but B doesn't, the sentence can still be seen as means for making B sit down. We are looking then, for entities not produced or designed by human beings, that display either intentionality, or teleology, and preferably both. If we understand what it is for such entities to be supposed to do something (have a function) or to be supposed to correspond to something, this might help us understand these notions as applied to linguistic expressions without direct appeal to the fact that they are artifacts.

Millikan's suggestion is that we start by looking at biological devices: hearts, kidneys, eyes, as well as reflexes such as eye-blinking, and instinctive behavior patterns such as mating displays. All these are said to have functions: a heart has the function of pumping blood, the eye-blinking reflex has the function of protecting the eye from objects that may enter it, the mating display has the function of attracting females, etc. The question then arises, where do these devices get their function from? What is it to say that it is those functions that they are *supposed* to perform? Millikan claims that the answer can be found in the evolutionary history of those devices: she defines the *direct proper function* of a device as that performance which earlier replicas of the device had performed, that accounted for their proliferation. It is that function which had a survival value for those devices in the past that had performed it, thus accounting for the reproduction of those devices. I will soon go in much greater detail into this definition. The crucial thing about it, however, is that it shows how things that are replicas of other things can be assigned with a function; or conversely, how the fact that things are attributed with functions can be explained without depending on teleological (or intentional) notions but instead just on their being replicated. If, then, we can similarly describe linguistic entities as devices that are being replicated, and that proliferate because of some function that they perform, we may be able to explain their teleological nature, and maybe even their intentionality. To do this, we need an analysis of the above notion of 'proper

function' that will apply to linguistic devices, as well as to biological ones. I will now describe Millikan's theory of proper functions, but will generally leave the description of how it applies to language until the full account has been given, except for some short remarks.

4.2 Theory of Proper Functions

4.2.1 Direct proper functions

The first thing that needs to be explained to make the above intuitive definition of direct proper function more precise, is the notion of replication. One device, B, is said to be a *reproduction* of another device, A, if (i) they have certain properties in common, (ii) their having these properties in common is explained by natural laws that relate B's having the relevant properties to A's having these properties, and (iii) those laws would correlate any changes in A in respect to these properties to identical changes in B, the direction of causation being directly from A to B. The common properties by reference to which B is a reproduction of A are called the *reproductively established properties* of B. Under this definition, genes are reproductions of other genes; similarly, a xerox copy of one document is a reproduction of another. But reproductions need not be thought of only as in these direct, "mechanical" cases. Any artifact that isn't an originally designed one but an imitation—say, a cross on a church wall—is also a reproduction in this sense. Imitated behaviors, such as shaking hands, or waving good-bye with a handkerchief, are also reproductions[2]. Given the assumption of a naturalistic account of the mind-brain, the psychological mechanisms that account for the imitation process will be, in a reductive account, natural laws

[2]In general, conventional behaviors are reproductions, as their being similar to one another is not explained by any intrinsic, functional relation between those behaviors and their outcome, nor are they produced by natural dispositions of the person performing them. Instead, they are produced as imitations of earlier such behaviors.

operating on a physical medium, the brain, that explain why the imitated devices are similar in certain respects to some original.

It is interesting to note here that Richard Dawkins, in his famous book 'The Selfish Gene' (1989), thinks similarly on the notion of reproduction. After a long discussion of genes as replicated elements, he argues that human culture is an environment with elements in it that are being replicated by imitation. He calls these elements 'memes', of which he says:

> Examples of memes are tunes, ideas, catch-phrases, clothes fashions, ways of making pots or building arches. Just as genes propagate themselves in the gene pool by leaping from body to body via sperms and eggs, so memes propagate themselves in the meme pool by leaping from brain to brain via a process which, in the broad sense, can be called imitation'.
>
> (*ibid*, p. 192)

Millikan points out that not all similar devices, biological as well as artificial, are reproductions in this strict sense. A mating display of one fish, for example, is not a reproduction of its father's mating display, because it is not a direct imitation but rather an instinctive behavior which is a product of its genes. The genes *are* reproductions of one another, but if one fish lost a fin and therefore would make an odd display, this wouldn't affect the form of its offspring's display. Similarly, mass-produced products are not reproductions of one another, but rather they look similar because they are produced by the same machine—they have certain causes in common. Language devices though (e.g. words, sentences, grammatical moods, intonation patterns, etc.), *are* reproduced items, Millikan claims. It is not that each sentence is a direct copy of an identical sentence, but rather different elements in the sentence are reproductions of other such elements. Thus, each word is a reproduction of earlier such words that the speaker has heard; and the grammatical form of the sentence is a reproduction of earlier such grammatical forms.

The fact that some items are direct reproductions of other items, while others are similar because of common underlying causes, makes for two kinds of classifications. A *first order reproductively established family* is defined as a set of entities having certain characteristics in common because of being reproductions of one another. A *higher order reproductively established family* is defined recursively on the basis of the first definition. It consists of entities having certain common characters because either (i) they are produced by members of a (first or higher order) reproductively established family having as their proper function to produce such items, or (ii) they are produced by the same device, the proper function of which is to produce such items that will match one another. Now, an *ancestor* of a member m of a reproductively established family is defined as either (i) any member of the same first order reproductively established family as m, where m was derived by reproduction (or successive reproductions) from it, or (ii) any temporally earlier member of the higher order reproductively established family to which m belongs that was produced by an ancestor of the device that had produced m, or (iii) any temporally earlier member of the higher order reproductively established family to which m belongs that was produced by the same device that had produced m.

The *direct proper function* of a device m is defined (roughly)[3] as (i) that function F that certain ancestors of m had performed, (ii) where the performance of F was causally connected to having a certain character that m has in common with its ancestors as members of the same reproductively established family, and (iii) the performance of F by some of m's ancestors (partly) explains the proliferation of that device family, and thus explains the fact that m exists. We see from this definition that not only direct reproductions can be assigned with direct proper functions, but that these can be assigned also to their products. If something that a replicated device produces accounts for its proliferation, it also accounts for the proliferation of this product. Thus, for example, if the fact

[3] I will not give Millikan's exact definition as this will require going into too much detail.

that certain genes are being replicated is explained by their contribution to the survival of the organism that carries them by producing a heart that pumps blood, the heart's proliferation can similarly be explained by its performance of that function.

It is interesting to explore some of the consequences of this historical-evolutionary definition of direct proper function. Notice, first, that the definition attributes to a device a function without depending on its actual performance. A device is said to have a specific function because it is a reproduction of some *other* devices that performed in a certain way. Thus, for example, a foetus may be born dead but nevertheless it does have a heart, kidneys and eyes with specific proper functions that they will never perform. Second, this definition of proper function determines which of the various things that a device does is its function, and which isn't. For example, hearts both pump blood and also make a ticking noise. But it is the first rather than the latter performance that is considered as the heart's proper function, as pumping blood is relevant to the proliferation of hearts, whereas making a ticking noise isn't. Accordingly, the proper function of a device is not dependent on how a device type functions on the average, or in most cases. A device may rarely perform its proper function, but the fact that it sometimes does perform in that way is what accounts for its proliferation. Sperm cells are an example that helps make this point: a sperm cell's proper function is to fertilize an ovum, even though only one in many million ever does perform this function. Nevertheless, it is the fact that they do, sometimes, fertilize an ovum that accounts for the proliferation of that device family. Finally, since the definition of direct proper function depends on earlier replicas rather than the performance of an actual device token, it follows that a device with no earlier replicas cannot have a direct proper function. As Millikan puts it, if a cosmic accident would result in producing something that was built, and performed, just like a human being, this creature would not have a heart, kidneys or eyes, because the devices in it that resembled ours would not have

the right kind of history. This creature cannot be said, for example, to have a heart failure because its heart is not supposed to do anything in particular, and thus cannot *fail* to do anything.

A further complication to this theory of proper functions arises from the fact that not every device that performs its proper function does so in a way that is really relevant to the explanation of why it proliferates. A detailed explanation, for example, of how a heart with a pacemaker attached to it performs its function of pumping blood doesn't really account for the proliferation of hearts in human beings. That is because the conditions to which most hearts have adapted didn't include pacemakers. Millikan therefore defines a notion of *Normal explanation*, which is an explanation of how members of a particular reproductively established family have in most cases performed, in the past, the particular proper function they have as members of that family. Part of such an explanation will have to mention the *Normal conditions* in which that device had performed, i.e. the external conditions that prevailed in most cases when such devices had actually performed their proper function. The reason 'Normal' is capitalized in this context is to distinguish it from 'normal' in the sense of a statistical average. A Normal condition for the performance, for example, of the eye-blinking reflex that is supposed to keep objects from entering the eye, is that there is an object that is about to enter the eye. This does not mean that the eye must be bombarded with such objects most of the time. It is only *when* a device performs its proper function (and this, we have seen, can happen quite rarely), that the Normal conditions must prevail, to make its performance accord with the Normal explanation for that performance.

Millikan distinguishes between different levels of proximity of a Normal explanation. The most proximate explanation is the least detailed one, which mentions only the minimal conditions for a device's Normal functioning, but will ignore, for example, how those conditions came about. An explanation of

how a device works may be Normal on the most proximate level, but be ab-Normal on a less proximate explanation. For example, the lungs of a person breathing under water with the aid of an aqua-lung are performing their proper function Normally on the most proximate level. They are not functioning Normally on a less proximate level, however, one that will also mention where the lungs usually get their oxygen supply when performing their proper function. As we shall see later, the most proximate Normal explanation determines the intentional object of devices that display intentionality.

4.2.2 Derived proper functions

Up until now, the notion of a proper function—that function that a device should perform—has been explicated in terms of its belonging to a reproductively established family. There are devices, however, that do not belong to any reproductively established family, but nevertheless we would want to say that there is something that they are supposed to do. Consider, for example, the brown skin of a chameleon placed on a brown surface. It seems plausible to say that this skin color has the function of making the chameleon invisible to predators, but the brown skin is not a member of a reproductively established family: it is obviously not a direct reproduction of other brown skins. But it is also not the proper function of the pigment-arranging mechanism of the chameleon to produce a skin color that is *brown.* Producing a brown skin is not something that accounts for the proliferation of that mechanism. Instead, its function is to produce a skin color that matches the color of the chameleon's surroundings, whatever color that may be. Now, if the brown skin of a chameleon doesn't belong to a reproductively established family, it cannot, by definition, have a direct proper function. Millikan explains in what sense this brown skin does nevertheless have the proper function of making the chameleon invisible to predators.

I have said above that the direct proper function of the pigment arranging mechanism of a chameleon is to produce a skin color that matches the color of its surroundings. As such, we may say that it has a *relational proper function*, because its proper function is defined in relation to something else. Now, given a specific color of its surroundings (call this given color the device's *adaptor*), the pigment arranging mechanism may be said to have the *adapted proper function* of producing that specific color. In our example, given a brown surrounding, the device has the adapted proper function of producing a brown skin. A device that bears the right relation to its adaptor (in our case the brown skin itself, in relation to its brown surroundings) is said to be an adapted device. We have explained above why the adapted device itself is not a member of a reproductively established family. But being a product of a device with a relational proper function, we may now say that the adapted device has, as its *derived proper function*, the same function as that of the device that has produced it. In our example, the brown skin has as its derived proper function to make the chameleon invisible to predators; this proper function was derived from the proper function of the pigment arranging device that has produced it.

The notions involved here are quite complicated, so I will try to show how they generalize by applying them to another example. Consider a bee-dance, that is supposed to have its shape related in a specific way to the relation between the sun, the hive, and a particular location of nectar that the bee has detected. The dance-choreographing mechanism in a bee thus has the general, relational proper function of producing dances, that by bearing a certain relation to their adaptors—the location of the nectar, sun and hive—will cause the bee family to move in a certain direction where nectar has been located. Now, given a specific set of adaptors, the dance-choreographing mechanism in the bee has the adapted proper function of producing a specific dance. This specific dance, being produced by this mechanism, will be an adapted device if it bears the correct relation to the specific adaptors. And, the specific dance will have the

derived, general proper function of moving the bee family in the direction of nectar (whatever that direction may be).

This example is in some sense different from the former example with the chameleon. The bee dance, apart from having a certain derived proper function, will also be a member of the higher order reproductively established family of bee-dances performed according to "B-mese" syntactic rules. As such, the dance will have as its *direct* proper function to move the bee family in a specific direction, because doing so is part of the explanation for the proliferation of bee-dances conforming to these syntactic rules. The bee dance, then, has its own, direct proper function, in addition to its derived proper function. Now, when the dance choreographing device performs its adapted proper function properly, it will produce a dance that will have as its derived proper function the same function as the dance's own, direct proper function. However, in a case of some failure of this device, it may produce a dance that will have divergent proper functions. Its derived proper function (that may fail to be performed) might be to move the bee-family in, say, the North-West direction, where the bee has located nectar; but the dance that has actually been performed, may have its own direct proper function of moving the bee-family to the South-West direction, instead.

We will turn to the discussion of human languages shortly, but I will now hint at the significance of this example to our later discussion: a linguistic expression may have a direct proper function—which comes close to its having a certain meaning—as a result of being a reproduction of several linguistic elements, each with their own direct proper function. At the same time, however, linguistic expressions, being products 'designed' by their utterers for a certain purpose, will have a *derived* proper function. This derived proper function (and meaning) will depend on the intentions of the speaker that uttered those linguistic expressions. In many cases the derived and direct proper functions of linguistic expressions

will converge, but in others (e.g. in metaphors) they may diverge.

Do devices with derived proper functions have a Normal explanation for their performance? Seemingly they shouldn't, because such devices may be new, so that there is no special way according to which they had previously performed. But in fact they do have such Normal explanations, although these will be quite general. Devices with derived proper functions perform these functions (e.g. making the chameleon invisible to predators, or moving the bee family in the direction of nectar) due to some relation they bear to their adaptors. A Normal explanation of those performances will explain how having such relations to the adaptors will have the effect it should have as a device with a certain derived proper function. In our examples, it will be a general explanation of how having colors identical to that of a chameleon's background makes it invisible to predators, or how a dance's having a certain form can determine the direction of flight for bees observing it.

As was mentioned in the introduction to this chapter, the notion of proper function is intended to be generally applicable. It should apply to any reproduced devices the proliferation of which is enhanced by some effects of their reproduced properties (not all reproduced items have a proper function, of course—your image in the mirror, for example, isn't 'for' anything). Some biological devices have been mentioned to which the notion applies. It also applies to different artifacts—tools, for example. It has been claimed that many of these are reproductions, and it is easy to see that they are being reproduced, retaining specific characters, because of the function that having these characters enabled them to perform. Hammers, for example, are not invented over and over again. Instead, they are copied, retaining their distinctive shape and weight, because these features enable them to perform certain functions that are found useful by their producers. They have, therefore, direct proper functions. We will now move on to see how the notion of direct proper function applies to other arti-

facts, that presumably are copied due to functions they perform, i.e. to linguistic devices.

4.3 Three aspects of meaning

Millikan aims to distinguish between what a linguistic device means, and what its utterer means. To make this distinction, she has to explain what it is for a linguistic device to have its *own* meaning. Her claim is that a linguistic device has its own meaning in the same sense as biological devices have their own function: there is something that language devices of a certain type are supposed to do, that accounts for their proliferation, for their being replicated over and over again. The problem of identifying a linguistic device's meaning, then, is that of identifying which of the several effects it might have are essential to its proliferation, and which are only accidental, or occasional uses it may be put to. This, she claims, captures one aspect of the notion of linguistic meaning, the teleological character mentioned above. How this is done will be described in the first of the following sections, that discusses the stabilizing proper function of language devices.

Another characteristic of meaning that needs to be captured, is intentionality. Millikan derives the explanation of this aspect from the first. A device is said to be *about* something, in so far as a certain correspondence relation between them holds when the device performs its proper function Normally. This rather vague claim will be explained in the second following section, which discusses the notion of 'sense', the second aspect of meaning.

Finally, Millikan discusses a third aspect of meaning, that of the intensions (with an 's') associated with referential terms. Intensions have to do with criteria of application for referential terms onto their referents that individual users

of language may use. This aspect of meaning, Millikan argues, is the least important one as it is basically not a public phenomenon. Intensions are not determinants of meaning, but are ways of *identifying* meanings that are determined by the proper function, and accordingly by the sense of a linguistic device. This aspect of meaning, intension, will be discussed in the third of the following sections.

It is interesting to note that Millikan's three aspects of meaning, i.e. proper function, sense, and intension, are respectively parallel to Hesse's definition of meaning as use, reference, and associated ideas. Millikan, however, is more explicit on what each of these elements consists of and how they relate; also, as we shall see, there is a big difference in the significance that Millikan and Hesse each assign to these various elements.

4.3.1 Stabilizing proper function as the first aspect of meaning

Stabilizing and standardizing proper functions

Like some biological devices, language devices are dependent for their functioning on some cooperative agents (or devices within these agents). A mating display, for example, with the proper function of attracting females, depends for its performance on a device in the female that will respond by being attracted to the performer of that display. Similarly, an imperative sentence, for example, that has the function of making a hearer act in a certain way[4], depends on a cooperative response from a hearer that will interpret and act according to what he has heard. This places two restrictions on the proper functions that language devices may acquire. First, the function of any linguistic device, if it is

[4]That this is its function will be argued for shortly.

to proliferate, must be one that will benefit both its speaker, so he will continue to produce it, and its hearer, so he will continue to respond to it. Second, since language devices, like mating displays, are used with many different cooperating partners, they must be standardized so that the same response can be expected from different cooperative hearers.

As a matter of fact, it is clear that not on all occasions the response of a hearer will actually benefit him: a speaker may be manipulative, or lie, or just be mistaken. A hearer that complied with the imperatives of such a manipulator, or believed a liar, might suffer because of being cooperative. But it is not such cases that account for the proliferation of the language devices involved: these proliferate *in spite* of such cases. And they proliferate because not all cases are such; instead, there must be enough cases where the hearers do benefit from their responses. We may say that these cases are the stabilizing ones: stabilizing both for the speaker who elicits the intended response, and for the hearer who benefits from responding in the expected way.

Again, not all actual cases of use of language devices are standard; a word may be used in a confused way (e.g. malapropisms), or a sentence may be used ironically. But again, these are not the rule—language devices proliferate in spite of such deviations, and in fact many such deviations are parasitic in their performance upon the standard. If we are, then, to identify the function that accounts for the proliferation of a linguistic device, we are looking for what Millikan calls its *stabilizing and standardizing direct proper function.* This is the function that a language device performs in enough cases, which keeps both speakers and hearers using the device and responding to it in a standard way, thus stabilizing its function.

Indicatives and imperatives

Let us consider, as an example, the devices that indicate the indicative and imperative grammatical moods. Millikan claims that their stabilizing and standardizing proper functions are the following: sentences in the indicative mood have the proper function of conveying information, i.e. of producing true beliefs in the hearers. Sentences in the imperative mood have the proper function of producing compliance. Obviously it is not the case that all indicative sentences do convey true information, or that all imperatives produce compliance; it is even doubtful if most of them do. But, as we've seen in the discussion of direct proper functions, this is not what matters, as claims about the proper function of a device are not claims about its invariant or average function. Instead, Millikan says that it is those functions of the indicative and the imperative mood indicators that 'keeps them going'; these language devices must function in that way in enough cases so they will still be produced by speakers and responded to by hearers.

Why *these* functions? Let us begin with the imperative mood. From the speaker's point of view, it is quite clear that compliance would encourage him to continue using imperatives; conversely, if his imperatives were *never* complied with, there would be no point for him to use the imperative form. It is not enough, notice, that imperatives should create intentions to comply; it is only the carrying out of the intentions that render the imperatives useful. From the hearer's point of view, we must assume that at least in some occasions he is reinforced, in various ways, for complying with imperatives; this is what keeps him complying. While the reinforcements may vary from one occasion to the other, what is common to all these cases is his *compliance* that is reinforced. Producing compliance, then, is the function of imperative sentences, that commonly reinforces both speakers and hearers to continue using that device. It is the stabilizing and standardizing proper function of the imperative mood indi-

cators.

Sentences in the indicative mood, Millikan claims, have as their stabilizing proper function to produce *true* beliefs. Speakers, we may assume, use indicative sentences to produce *some* beliefs in hearers. They are rewarded in various ways by their hearers, however, for speaking the truth. Hearers, again, benefit mainly because of the true beliefs they acquire from speakers. Conversely, if no true beliefs ever resulted from hearer interpretations of indicative sentences, hearers would cease to respond to them by creating beliefs, and so eventually speakers would stop producing them. We may conclude that producing true beliefs, or conveying information, is the stabilizing proper function of the indicative mood indicators.

The claims made, that the imperative and indicative moods function in the way described above, are not really new. But the advantage of the approach offered by Millikan is that it tells us what we are looking for in the functioning of language devices, i.e. their stabilizing and standardizing proper functions. This gives a ground for making stronger claims than just generalizations made on the basis of a statistical distribution. Also, it explains why and how some effects that language devices have can be distinguished as their proper function, while others can be ignored as irrelevant. It is a well known puzzle whether linguistic claims are only descriptive, or whether they have a normative force, that can distinguish between 'right' and 'wrong' uses of language. Millikan shows how descriptive claims of the right kind can have at the same time the required normative force[5].

Notice that the imperative and indicative grammatical moods are not names of reproductively established families. Not all mood indicators are copies of

[5]To avoid misunderstanding: of course it is not claimed that using a linguistic device in a way that does not conform with its proper function is wrong in any moral sense. The norms involved here are morally neutral, in the same way as saying that cooperation is good for an ant colony has no moral significance as to how ants should behave.

one another, as is evident from the fact that they have different realizations in different languages. Instead, it is a classification according to the common proper functions that different reproductively established families of language devices have. Each particular device has the proper function it does as a reproduction of other devices that are identical to it in form, not only in function.

The proper function of particular sentences

The proper function of the grammatical moods is quiet general—the indicative is to convey information, the imperative to produce compliance. But *what* information should a particular indicative sentence convey? And which act will count as compliance with a specific imperative? This is determined by the proper function of other parts of the sentence, other language devices that the sentence is a reproduction of. As was explained earlier, sentences are never as a whole copies of other exactly such sentences (except, perhaps, idioms). Instead they are reproductions of various devices, that function together. Each of these devices is dependent for its functioning, as a Normal but also a necessary condition, on the functioning of all other devices in the sentence. Some of the devices determine the propositional content of a sentence, but this alone doesn't determine its function: the function depends also on the grammatical mood indicators, that determine what effect that has to do with this propositional content the sentence should bring about.

An important class of language devices that are relevant to the determination of the propositional content of sentences are the denotative and referential ones, and their proper function, Millikan claims, is to precipitate acts of identification of variants in the world to which they correspond[6]. I will not go into

[6]Denotative and referential language devices include both predicates and names, which correspond to variant aspects of the world. Millikan rejects an ontological distinction between the categories of 'substance' and 'properties' as mutually exclusive categories.

an explanation of why this is their proper function, and I will not describe the proper function of other language devices—these may be found in Millikan's book. What is important, however, about the denotative and referential terms is not so much their proper function, but the fact that they are intentional devices. Intentional devices are classified as such not according to any proper functions they may have in common; instead, intentionality has to do with the *way* that intentional devices carry out their proper functions. This will be explained in the next section, that discusses the second aspect of meaning, intentionality.

Stabilizing proper functions and user's intentions and purposes

The first aspect of meaning, i.e. the stabilizing proper function of language devices, has been defined such that it is clear how meaning is independent of the intentions of a particular user of a linguistic device: a linguistic device's direct proper function is determined by the performance of its ancestral devices, not by its producer's intentions. The direct proper functions of language devices, however, are not independent of speakers' intentions in general. The survival of a public linguistic device without change of function depends on there being a critical mass of occasions where speakers and hearers use that device such that it performs its stabilizing function in accordance with the speakers' and hearers' purposes. As was mentioned earlier, a linguistic device has, apart from its direct proper function, a derived proper function, derived from the speaker's and hearer's intentions or purposes in using it. In the Normal (and stabilizing) cases, its derived proper function will coincide with the linguistic device's direct proper function.

On the basis of the notion of a language device's direct proper function, we can now see in what cases language devices will fail to perform their direct proper function. It is important that we distinguish these cases where a linguistic

device fails to perform its direct proper function, as claims we shall make about its functioning need only apply to the cases where it is functioning properly. In particular, claims that will be made about the mapping relations that hold between elements of a sentence and aspects of the world will only apply to sentences that are constructed of properly functioning language devices. This is analogous to claims made about the functioning of biological devices: our description of how, say, a heart functions need only apply to properly functioning hearts, not to *all* hearts, e.g. dead or diseased ones.

As I've explained, the proper functioning of language devices depends both on their producer's purposes and on their interpreter's purposes. Accordingly, failures in their functioning can be divided into cases of failure in the functioning, or deviant intended uses, of either the speaker, or the hearer, or both. Failures in the functioning, for example, of the indicative and imperative mood indicators will be the following. (1) Cases in which the speaker purposes the stabilizing function but the hearer does not: this could happen if the hearer doesn't understand the speaker, or refuses to cooperate for some reason. (2) Cases in which the hearer purposes the stabilizing function, but the speaker does not: the most significant example is lying. The hearer takes what has been said as true information, the speaker intends that he does so, but the speaker does not intend his utterance to produce a true belief. (3) Cases in which both speaker and hearer do not purpose the stabilizing function: these are cases of parasitic use, such as indirect speech acts, sarcasm, and metaphor[7]. In all these cases speaker and hearer may manage to understand one another, but their understanding in those cases cannot account for the proliferation of the devices being used. Like the case of lying, not all language can be metaphorical, for example. That people communicate using metaphors depends on the fact that their metaphorical expressions have stable and standard uses that are other

[7]Millikan, we see, treats metaphor as a parasitic use of language devices, which is divergent from their stabilizing function. In chapter 7, I will claim that metaphors do have a proper function, apart from that which is used parasitically.

than their improvised, metaphorical uses. (4) Cases in which both speaker and hearer purpose the standardizing function, but this function is not performed: these may occur when the hearer tries, but fails to comply with an imperative sentence, or a speaker is mistaken about the facts that he conveys to his hearer.

Purposes vs. intentions

It is crucial to understand what this 'purposing' on part of the speakers and hearers means. Such purposing does *not* necessarily involve any explicit intentions that speakers and hearers should have. All that is required for a linguistic device to perform its stabilizing function Normally is that it will be identical with the speaker's utterance derived proper function, and with the derived proper function of the hearer's response mechanism, given that these derived proper functions have been performed. It is helpful to think of the example of the bee-dance, to understand this point. We need not assume that bees have any such thing as explicit intentions when performing their dances, or when responding to them; we may well think of these as automatic processes. It is true that peoples' use of language is different from bee-dances in at least the respect that the latter involve only inborn mechanisms, while the former are learned behaviors. But learned behaviors may still be described as acquired automated responses, rather than explicitly and rationally calculated ones. Millikan claims that many of our linguistic exchanges, in particular those where language devices perform their direct proper functions, are more accurately described in terms of such automated mechanisms, rather than by explicit intentions and rational calculations over such intentions, as Grice and his followers would have it. In Millikan's words: 'In the most usual cases understanding speech is a form of direct perception of whatever the speech is *about*. Interpreting speech does not require making any inferences or having any beliefs even about words, let alone about speaker intentions' (p. 62). The point made here is quite general. 'Any act that

is purposive has, as such, a proper function, but relatively few of our purposive acts... are backed by explicit intentions' (p. 62). This applies to *derived* proper functions of language devices, even though these are dependent on the speaker's purposes in communication. Millikan goes into lengthy argumentation to defend these claims, which I will have to omit here. It is important, however, that the reader should realize what is meant, in her account, by the notion of purposes that underlie the derived proper functions that language devices have when being uttered.

It is interesting that even Bennett, who generally follows and accepts the gist of Grice's analysis of meaning as dependent on intentions to communicate, has his doubts as to what role these intentions really play in communication (cf. Bennett 1976). He defines what he calls 'sub-Gricean conditions for meaning', which drop Grice's demand (cf. Grice 1957) that for an utterance S to mean P, the utterer U must intend that his hearer will recognize his intention to communicate P by uttering S. Accordingly, he argues that for a hearer to come to believe P he only needs to make the connection that when S is uttered P is true, and he need not consider that U has intended to make him believe that P (Bennett 1976, pp. 171–5). Later on he says that while he doesn't doubt that human communication largely does involve the Gricean mechanism of hearers considering speakers' intentions (*ibid*, p. 200), it is very difficult to find direct evidence for that (*ibid*, p. 201). In other words, he thinks that a hearer's response to S of coming to believe that P would be rational even if he didn't consider U's intentions. Bennett's considerations suggest that it is not a coincidence that it is difficult to find such direct evidence: perhaps, as Millikan argues, many cases of communication do not actually involve any such complicated intentions and calculations on part of the speakers and hearers. Speakers just talk and hearers just react, and only when invited to contemplate their reasons for talking and reacting would they actually consider the underlying purposes as explicit intentions.

Despite of what had just been said, it should be made clear that Millikan is not arguing that explicit intentions and calculations are *never* involved in communication. These do have a role when speakers and hearers are using improvised signs that have no direct proper functions (because of not being replications), or are using language devices in parasitic ways, such that what they are trying to do must be deduced, rather than directly perceived, from what is being said. It is exactly these cases where we cannot say that the speaker's *utterance*, or whatever sign he uses, means something true or false, but can only attribute truth and falsity to the speaker's intended message.

4.3.2 Sense as the second aspect of meaning

.

The first characteristic of meaning, that was discussed above, is that of meaningful language devices' being instruments for *doing* things. The second characteristic of many meaningful linguistic elements is their intentionality, i.e. their being about other things. Millikan sets out to give a naturalistic account of what it is for words and sentences to be about aspects of the world. One important thing to notice about intentionality, is that it cannot be expressed in terms of (proper) function. That one thing is about another is not a *function* of that thing, because being about something is not doing anything. Accordingly, a device's being about something cannot itself account for its proliferation, as this aboutness alone doesn't have any effects that can be or not be advantageous. Intentionality, instead, will be explained in terms of the conditions that must hold for a device to Normally perform some other proper function it has.

One feature of the notion of intentionality that needs to be captured by any account, is that intentional entities can be about things that do not exist, or that are not actual. Millikan aims at explaining this feature by the notions

of properness, or Normality. An intentional device is one that is *supposed to* correspond to something, in some sense of 'being supposed to' that needs to be explained. Being supposed to be about X allows, of course, that X is absent.

Millikan, as we shall see, explains the intentional aspect of meaningful language devices in terms of correspondence, or mapping relations that are supposed to hold between a linguistic device and some aspect of the world. The problem with explaining correspondence, or mapping, is that any set of things can map onto any other set, according to *some* mathematical mapping function, or in fact any number of mathematical functions. There is no logical priority to one mapping over another. Millikan will have to show, therefore, how a specific mapping relation—the one that holds between a representation and what is represented—is somehow special and different from other possible mapping relations.

Intentional icons

Let us see now how Millikan answers these difficulties and objectives. She starts out by defining a technical term, 'intentional icon', that applies firstly to sentences, but also to other kinds of devices that will not be discussed. Intentional icons are devices that are supposed to map in a certain way onto the world, in order to serve their direct proper function Normally; that is, in most cases they do so map when serving their function. Also, they are devices that are supposed to be used or interpreted by cooperating devices. We shall look, specifically, on imperative and indicative sentences as intentional icons.

The first characteristic they have, as intentional icons, is that each particular sentence is a member of a reproductively established family with a direct proper function. A sentence as a whole belongs to a reproductively established family

of sentences displaying the same overall syntactic form. As such, the sentence has a relational character: it is Normally constructed of other elements, each having a specific *kind* of proper function (e.g. it is constructed of nouns, verbs, etc. arranged in a certain order). The sentence as a whole is then adapted (in the above technical sense, discussed in the section concerning relational proper functions) to the specific words in it, each with its own direct proper function. This is what gives the sentence a specific, adapted proper function. As was mentioned earlier, each device within the sentence (including its syntactic surface form) depends for its Normal functioning on the presence of the other devices.

The second characteristic of indicative and imperative sentences that makes them intentional icons is the fact that they stand midway between a producing device, and an interpreting device, that are standardized to fit one another, and have the presence of each other as a Normal condition for their proper performance. In other words, sentences are the medium through which the standardized cooperating devices operate. The sentences' proper functions, as their 'reason for being', is to be explained in terms of the contribution they have to the coordinated functioning of those other two devices.

Thirdly, indicative and imperative sentences adapt the cooperative, interpreting device to conditions that will enable it to perform its proper functions. It is here where the difference between the indicative and the imperative becomes apparent. The interpreter device's proper response to an indicative sentence is to create a belief, and in the Normal case this belief will be true. The Normal explanation for *why* this belief is true makes reference to the fact that the sentence mapped onto the world in a certain way. It is in this sense that the indicative sentence adapts the interpreting device: it gives the device's relational proper function of creating true beliefs a specific content, making it have an adapted proper function, to create some specific belief. The belief that is created will itself bear specific mapping relations to the adaptors of the indicative

sentence that adapted it, i.e. to the conditions that made it true. The interpreter device's proper response to an imperative sentence, on the other hand, is to produce certain conditions that will bear specific mapping relations to the sentence. Again, the sentence adapts the interpreter device's relational proper function of complying, to a specific content. It then has a specific, adapted proper function of producing certain conditions that will map onto the imperative sentence according to a particular mapping function.

Millikan provides a useful example that helps clarify the difference between indicative and imperative icons, as well as explaining what an intentional icon is about. Consider the pattern of impulses, P, that passes through some particular cross section of an animal's optic nerve. P Normally maps (1) onto a pattern of retinal stimulations that caused P, (2) onto a pattern of optic nerve outputs to the brain that P causes, and (3) onto various patterns occurring in other cross sections of the optic nerve, some of which caused P, and some that are caused by P. P is an *imperative intentional icon* of the *last* member of the series of things that it is supposed to map onto and produce. P is an *indicative intentional icon* of whatever it maps onto that must be mentioned in giving a Normal explanation for the proper performance of its interpreting device, as adapted to the icon P.

P is thus an imperative intentional icon of a certain aspect of the visual percept that the optic nerve produces rather than of any of the other impulses it causes along the optic nerve. Similarly, an imperative sentence is an intentional icon of the conditions that the hearer should produce, rather than, say, the intentions it Normally causes that also bear a mapping relation to the sentence. P is an indicative intentional icon of some aspect of the world about which it carries information, rather than, say, the retinal pattern, because to perform its proper function it must must carry information about the world (a stimulus pattern of an optic nerve cross section caused by an optic illusion cannot perform its proper function of conveying information). Similarly, an indicative sentence

performs its proper function of creating a true belief because of its mapping onto the world, rather than, say, mapping onto the intentions or beliefs of the hearer that produced the sentence.

Real value and sense

Millikan calls the aspect of the world that a properly performing intentional icon actually maps onto in the Normal case, its *real value*. The mapping function according to which it is *supposed to* map onto a real value is an intentional icon's *sense*. Thus, imperative sentences that are not complied with do not have a real value. Similarly, indicative sentences that are false do not have a real value. Both such indicatives and imperatives, though, have a sense. I said earlier that Millikan explains intentionality in terms of the notion of 'being supposed to' map; we see now that her notion of the 'sense' of an intentional icon is what explains this idea of being supposed to map in a certain way. The sense of an expression is that mapping rule it accords with when it performs its proper function Normally. From the explanation of what it means for a linguistic device to be supposed to perform its proper function, we derive the notion of its being supposed to map in a certain way: this mapping usually underlies its performance of the function it is supposed to perform.

As was mentioned earlier, it is only full sentences, rather than separate words, that have actual, specific proper functions. Words and syntactic forms have only relational proper functions, and must be embedded in the context of a sentence to have adapted proper functions. The most basic unit to have proper functions, and therefore to have a real value and a sense, is the sentence. *Within the context of a properly functioning sentence*, a specific referential term will have a real value—its referent. A lone term, or one in the context of, say, a false sentence, doesn't have a real value; there is nothing it corresponds to. It only has a sense,

something that it is supposed to correspond to, when in the context of a true (or obeyed) sentence. That a term has a sense, therefore, does not imply that it has a real value.

This way of presenting things implies that there is an important difference between the way sentences correspond to the world, and the way referential terms do. From the fact that a sentence is supposed to correspond to something—some world affair—it does not follow that there is any such world affair. But from the fact that a referential term is supposed to correspond to something, it does follow that the thing to which it is supposed to correspond really exists. That is because the referential term's being supposed to correspond to something depends on its reproductively established family having a history that involved actual correspondences with the referent, in the context of properly performing sentences. Conversely, a vacuous term such as 'Pegasus', is not supposed to correspond to anything, because it never did actually correspond to something within the context of a properly functioning sentence (e.g. within a true indicative sentence). 'Pegasus' does not have a sense.

To summarize this point: the sense of an intentional icon (e.g. an indicative or an imperative sentence), or of one of its elements (e.g. a referential term), is its having as a Normal condition for the performance of its direct proper function that it maps onto something else according to some specific mapping rules. The sense of a language element is its Normal mapping rule.

It is important to notice the difference between Millikan's notion of sense, and the notion of intension (with an 's'). Sense is just a mapping rule, whereas intension is a criterion for the application of a term, a criterion known and used by language users. A term's intension has to do with speaker's dispositions in applying a term, or the justification in terms of beliefs, or sensory stimulations, that they have for applying it. We will turn now to discuss this notion further,

but the important thing to understand is how it is different from a term's sense. A term's sense is only the mapping function according to which it should correspond, and it doesn't matter how this correspondence came about. The sense of a referential linguistic device is stable and common to all cases of its use. The intensions associated with referential terms, we will shortly see, may vary to a large extent between different users, and in different occasions of use. The difference between sense and intension is apparent in cases of terms that have one but not the other. 'Pegasus', for example, has an intension, because people have descriptions of it. But it cannot have a sense, we have seen, because there could be no sentence which performed its proper function due to a 'Pegasus' token mapping onto something real, since it is a vacuous term.

4.3.3 Intension as the third aspect of meaning

The first aspect of meaning of any public-language device is its stabilizing function. Denotative and referential terms have also another aspect to their meaning, their sense. We now come to another aspect of the meaning of referential terms, their *intension.* Intension, we shall see, is problematic for an account of the public aspects of language, as it is basically not a public phenomenon. If we compare language devices with tools, both having their own proper functions, the intension is analogous to the particular know-how that each user of a tool develops, as to how he handles that tool. Even though both tools and language devices have stable functions, each user may employ them somewhat differently, to achieve the same result.

For speakers to apply a term, they use a certain concept of it that they may have; the concept that governs their use of a term is what Millikan calls its intension. She divides intensions into explicit and implicit ones. Explicit intensions correspond to definite descriptions of a term, such as 'a yellow citrus fruit' for

lemon, or 'Britain's prime minister' for Mrs. Thatcher. Implicit intensions are methods for recognizing things directly from perceptual data, and may not be expressed by using other words. There are, for example, many different ways of recognizing a lemon, each being a separate intension: it can be recognized by its looks, its taste, its smell, etc. The intension, i.e. the reason or technique that one employs to identify the reference of a term, may vary for an individual, between different contexts of his use of a term. The intensions for a term will also vary between individuals, and there is no reason to suppose that all speakers and hearers in a language community will share the same intensions for any term.

Is there, however, any sense in which there is a particular intension that is "the" intension of a public language device? Is there an intension that a language user must grasp, if he is to understand a referential language device? Millikan argues that there are no such public intensions that a language user must know in order to use a public term correctly. This is what enables one, without paradox, to use a term X that he has no idea what it means in sentences such as 'could you tell me what X is ?'. Millikan argues that this is really a question about X, the referent, rather than about 'X', the sign. I will not, however, go into her argumentation for making this point.

This claim about varying intensions, however, does not eliminate the possibility of defining a public intension for some terms (though not for all). A public intension for a term may be defined as some averaging over those intensions that people find most reliable, or in some cases those intensions that experts would find most reliable. So defined, the public intension of a term might be said to be one kind of 'meaning' it has, and this fits the intuitions of various philosophers as to what meaning is. In the case of proper names of people, it could be expected that there will be minimal, or no overlap among the intensions of all competent users of these names, as each user has his own ways of recognizing

the person to whom a name applies. In the case of such proper names there are no experts to turn to. Though proper names do have proper functions and sense, they do not have a public intension in the way it was defined. This fits with intuitions of some that deny that proper names have meaning.

On the other extreme, there are names of natural kinds that are studied by experts. There may be large agreement among those experts as to which are the most reliable intensions for such terms, and as such these terms may be said to have quite definite public intensions. Another case is that of things that are familiar to almost everyone, in the sense that everyone is in command of the core of highly reliable means for identifying these things. Examples would be words like 'apple', 'red', and 'square'. Such terms again can be said to have public intensions, and these will tend to be implicit, rather than explicit intensions.

As I've mentioned earlier, this analysis of meaning in terms of intensions is quite close to Hesse's notions of associated ideas that are part of the meaning of linguistic terms. The difference, of course, is in the significance that Millikan and Hesse each gives to this kind of analysis. Hesse thinks the associated ideas are an inseparable part of the meaning of any linguistic term, and thus sees meaning as quite variant, and to a large extent intractable as a stable phenomenon. Millikan agrees that in such an analysis meaning is in fact varied and unstable, but argues that for this reason it is the least important aspect of meaning. Stabilizing proper functions, and sense, are the stable, generalizable notions that should replace intensions in an adequate account of public meaning.

Millikan's position on the role of intensions in language can be summarized as follows. Intensions, rather than being infallible determinants of meaning, are fallible approximations of it. The sense of an expression is fixed independently of its associated intensions, and those will, under various conditions, work better or worse for identifying the referents that are determined by the sense. That

is what Millikan means by her talk of intensions being more or less reliable—something that couldn't make sense if intensions were to determine reference.

The ideas presented in this chapter are quite difficult to grasp, especially in the unavoidably compressed way they have been described here. Moreover, even when understood, many of the ideas are quite radical and open to objection. In the next chapter I will discuss in detail some of the objections raised by Fodor against Millikan's theory. I hope that this will both deepen the understanding of her theory, through the discussion, and answer certain interesting objections that can be devised.

Chapter 5

A Reply to Fodor: Who Really Solves the Disjunction Problem?

5.1 Introduction

Fodor and Millikan are both looking for a naturalistic theory of intentional content, or meaning (Fodor unpublished, and Millikan 1984)[1]. That is, one which will specify sufficient conditions for intentional content, or meaning, in terms that are not themselves intentional or semantic. Fodor provides such a theory that specifies the sufficient conditions that make a symbol have a certain meaning. He argues that his theory manages to solve a crucial problem ('the disjunction problem', to be presented later on) that other naturalistic theories of meaning, in particular those that appeal to teleological explanations as Millikan does, fail to solve. This chapter is intended to defend Millikan's position from Fodor's criticism.

In the first section I will argue that in a large number of cases Millikan's conditions for meaning satisfy Fodor's own conditions. Fodor, who doesn't notice this overlap between their theories, criticizes Millikan's theory as a whole.

[1] All page references in this chapter, unless otherwise stated, are to Fodor's paper.

His criticism applies, therefore, also to those cases that do satisfy his own conditions, rendering his position inconsistent. In the second section the disjunction problem will be presented, together with Fodor's argument for why teleological accounts fail to solve it, and how his account does. Next, I will present some problems with Fodor's solution. In particular, I will show how the conditions for meaning he suggests give conflicting results as to what the content of an intentional state is, in an example he considers. In the following section I will analyze what gives rise to Fodor's apparent inconsistency in his criticism of Millikan's theory, and try to reveal what the underlying differences between their approaches really are. Finally, I will present a problem which is intended to show that it is Fodor's theory that fails to withstand the acid test he has presented naturalistic theories with, i.e. it fails to solve the disjunction problem. Millikan's theory, I argue, is not affected by the problem I present.

5.2 Two naturalistic theories of meaning

Fodor suggests the following two conditions as sufficient for attributing to a symbol a specific intentional content: a symbol type 'X' means the property X if

1. (some) Tokens of 'X' are caused by X.
2. Tokens of type 'X' that are caused by things other than X, are asymmetrically dependent on (1).

His example:

> 'Cows cause 'cow' tokens, and (lets suppose) cats cause 'cow' tokens. But 'cow' means *cow* and not *cat* or *cow or cat* because

> there being cat-caused 'cow' tokens depends on there being cow-caused 'cow' tokens, but not the other way around. 'Cow' means *cow* because... noncow-caused 'cow' tokens are *asymmetrically dependent* upon cow-caused 'cow' tokens'.
>
> (p. 38)

The intuition behind this proposal is the following: let us assume that a symbol token of 'cow' (whether it is a 'Mentalese' token, or a public language one) means *cow*. If we can capture this meaning relation in terms of only *causal* relations between cows and 'cow', we will have achieved the naturalistic goal of characterizing meaning in non-intentional, non-semantic terms. What stops us from doing so, however, is that 'cow' tokens are caused not only by cows, but also by other things, e.g. cats. So, we must try to find something about the causal relation between 'cow' and cows that is different from the causal relation between 'cow' and cats. We must, in other words, explain how *error* is possible.

Fodor's suggestion is that the difference lies in the dependency between the causes of 'cow'. A cat causes 'cow' only because cows do; if cows wouldn't cause 'cow', neither would cats. But the fact that cows cause 'cow' is not so dependent on cats (or on anything else) causing 'cow'; cows would continue to cause 'cow' even if cats didn't. The dependence between the causes is an empirical fact, just like the causes themselves. This suggestion, then, if correct, characterizes meaning in naturalistic terms, as required.

Millikan has also tried her luck (and skill) in suggesting a naturalistic account of intentionality and meaning. Very roughly, what she proposed was the following[2]: a symbol type 'X' means X if

1. Tokens of 'X' must Normally map onto Xs.

[2]This way of presenting Millikan's ideas is very remote from the way she presents them. It is my adaptation of these ideas using terms as close as possible to Fodor's, to enable a comparison between the two theories.

2. Tokens of type 'X' that are caused by things other than Xs, are asymmetrically dependent on (1).

Millikan's first condition, in a less telegraphic version, means the following: devices of type 'X' that perform their proper function, co-occur in most cases with X when performing it. Thus, a Normal explanation for their performance will have to mention, as a Normal condition for their performance, the fact that they map onto (or co-occur with) X. The first condition's allusion to the notion of Normality assumes, of course, that tokens of 'X' have a proper function, and accordingly that they belong to a reproductively established family that has this proper function as part of the explanation for its proliferation.

Notice, now, that in many cases tokens of 'X' that satisfy Millikan's first condition, also satisfy Fodor's first condition: Xs that cause 'X' are a special case of a mapping relation between 'X' and X. If it is the case that the fact that X causes 'X' is what enables 'X' to perform its proper function, then it follows that 'X' will have to co-occur with X in most cases where 'X' performs its proper function, and will thus Normally map onto it. It should be noted that this does not mean that 'X's *do* map, or *are* caused by X in most cases; it only means that on most cases *where 'X' performs its proper function* it does so map onto X, but 'X' only needs to perform its proper function on some (and enough) occasions.

A few words of explanation for why Millikan does not require that X will be the *cause* of 'X' but only that 'X' will map onto X. The most obvious case where 'X' means X without being caused by X, according to her theory, is that of imperative intentional icons. In these cases the symbol-device 'X' temporally precedes the production of its real value X, so X could not be the cause for 'X' (in a sense, it is the symbol 'X' that causes its real value which it maps onto). But even in the case of indicative signs it need not be the case that a symbol-

device is caused by what it means. Millikan gives the following example where what an indicative sign means isn't one of its causes:

> '...certain species of northern hemisphere bacteria...orient themselves away from toxic oxygen rich surface water by attending to their magnetosomes, tiny inner magnets, which pull toward the magnetic north pole, hence pull down...The function of the magnetosome thus appears to be to affect that the bacterium moves into oxygen-free water. Correlatively, intuition tells us that what the pull of the magnetosome represents is the whereabouts of oxygen-free water. The direction of oxygen-free water is not, however, a factor in *causing* the direction of pull of the magnetosome.'...'What the magnetosome represents then is univocal; it represents only the direction of oxygen-free water'.
>
> (Millikan 1989, pp. 290–91)

Let us concentrate on those cases which satisfy both Millikan's requirement for 'X' to Normally map onto X, and Fodor's requirement that X causes 'X'. I would like to point out that those cases also satisfy Fodor's second condition (and therefore also Millikan's, formally equivalent, second condition): if anything other than X causes 'X', its causing of 'X' asymmetrically depends on X causing 'X'. This is why: the proliferation of devices of type 'X' depend on the performance of a certain proper function by some members of the 'X' reproductively established family. In particular, the fact that a specific token of 'X' is produced, depends on the fact that enough members of the reproductively established family to which 'X' belongs have performed their proper function. That performance of each member, in turn, depends on the Normal mapping conditions that must hold when it performs its proper function. We may thus say that the production of a specific 'X' token, whether it is caused by X or by a non-X, depends on the Normal mapping of 'X' onto X on enough occasions. In those particular cases where the mapping relation is also a causal one (cases that would satisfy both Fodor's and Millikan's first condition), the production of X, if it is caused by a non-X (or even by an X), is dependent on enough 'X's

being caused by X. Moreover, the production of 'X' is *asymmetrically* dependent on 'X's being caused by (and mapping onto) X, as earlier productions of 'X' (by a device token's ancestors) cannot depend on it. The latter, of course, is Fodor's second condition for meaning, and we may therefore conclude that a symbol which satisfies both Millikan's and Fodor's first condition for meaning, *ipso facto* satisfies Fodor's second condition.

Let me repeat now Fodor's example of the symbol 'cow', and see how Millikan's conditions apply to it.

> 'Cows cause 'cow' tokens, and (lets suppose) cats cause 'cow' tokens. But 'cow' means *cow* and not *cat* or *cow or cat* because there being cat-caused 'cow' tokens depends on there being cow-caused 'cow' tokens, but not the other way around. 'Cow' means *cow* because... noncow-caused 'cow' tokens are *asymmetrically dependent* upon cow-caused 'cow' tokens'.
>
> (p. 38)

Millikan's conditions, it turns out, account for the same story. Think of a tribe of cow-hunters, that use the exclamation 'cow!' to alert their fellows to the presence of a cow. For them, an exclamation 'cow!' when cows are around will be useful, and therefore will be repeated. The same exclamation when only cats are around will not be useful—it might in fact be distracting; and the same exclamation when cats or cows are around will be useful only when cows are around. When we compare the two accounts we seem to get a rare occasion of agreement between two philosophers: the fact that cat-caused 'cow' tokens are produced is asymmetrically dependent on cow-caused 'cow' tokens in the required way. If 'cow' tokens do proliferate it is because there are enough cases of their being caused by cows, and they proliferate *in spite* of the fact that sometimes they are caused by things other than cows. 'Cow' tokens would proliferate if they were produced only when cows were around, but would die out

if they were produced only by mistake[3]. The mistaken cases, such as cat-caused 'cow' tokens, depend for their production on a sufficient level of cow-caused 'cow' tokens, but not the other way around. It is for this reason that 'cow' means *cow* rather than *cat*.

But, alas, things are not that simple. Fodor thinks that Millikan's account is wrong, and his is right. From comparing the two accounts we saw that in one sense Millikan demands more than Fodor does, in that she specifies that what sustains the asymmetric dependence should be evolutionary considerations, not just general causal dependencies. In another sense, Millikan's account is more permissive than Fodor's, in that she only requires mapping relations to hold between a symbol and its meaning, where Fodor requires causal ones. Fodor may think her account is wrong then, because it is unnecessarily restrictive in some respects, or that it is too permissive in requiring only mapping relations where causal ones are needed[4]. However, this is not what Fodor argues. Instead, he criticizes Millikan's account for not solving a problem (the disjunction problem, to be explained shortly) that his account does solve by introducing the asymmetric dependence condition. If her account does satisfy *that* condition, as I claim it does, and it is the asymmetric dependence condition that solves the disjunction problem, Fodor must be inconsistent in his criticism of Millikan. In what follows I will try to explain what gives rise to this inconsistency. But first, let us turn to see what the disjunction problem is, and why Fodor thinks that Millikan's account fails to solve it while his account succeeds.

[3]Notice that even if there was a close correlation between 'cow' tokenings and cats, 'cow' wouldn't mean *cat* because what we're after is not just a correlation. In the case described we assumed that 'cow' has a proper function that has to do with cows, and if referring to cats doesn't have a selectional advantage, any amount of correlation between 'cow' and cats wouldn't account for the proliferation of 'cow' tokens.

[4]In fact, Fodor himself is not too committed to *causal* relations between 'X' and X, and doesn't a-priori have anything against theories that wouldn't employ them. As he says, 'The treatment of error I've proposed is, in a certain sense, purely formal. It invokes a dependence among relations, and it is *compatible* with assuming that the relations among which this dependence holds are causal... But strictly speaking, it doesn't *require* that assumption' (Fodor 1987, p. 110).

5.3 The Disjunction Problem

The main defect that Fodor sees in teleological accounts of meaning is their inability to deal with what he calls *the disjunction problem.* He addresses this criticism to teleological accounts in general; my response will only consider its application to Millikan's version. If my defense of her theory proves to be successful, however, it will of course imply that the flaw is not in teleological theories as such, but perhaps only in specific formulations.

The disjunction problem is the following: suppose that 'X' means X if it is a law that Xs cause 'X's. Now there are two possibilities. For example: (i) Only dogs cause 'dog's. But then, all tokens of 'dog' must be true, because they will never be caused by anything other than a dog. (ii) Some non-dogs cause 'dog's. For example, both dogs and cats-on-a-dark-night are sufficient to cause a 'dog' token. From this it follows that 'dog' means *dog or cat-on-a-dark-night.* Thus 'dog' will be true both of dogs and of cats-on-a-dark-night. Our problem is that in either case we cannot make sense of what it is for 'dog' to be false, because 'dog' will never be tokened in the wrong circumstances. Another way to look at it: assuming (as we must) that errors do occur, we cannot distinguish between cases where a disjunctive concept is correctly applied, and cases where a non-disjunctive concept is wrongly applied. Suppose we stick to this naturalistic strategy of trying to define meaning in terms of causation, and retain the condition that for 'X' to mean X it must be the case that X causes 'X'. What the above seems to indicate is that this condition alone is not sufficient for a definition of meaning. What else is needed?

A general form of solution, Fodor says, is to distinguish between two types of situations. A type 1 situation is one in which whatever causes 'X' is in its extension; type 1 is a situation where the meaning of a term is constituted. A type 2 situation is any other situation, where anything can cause 'X', but 'X' is

true only of things that caused 'X' in the type 1 situation, and false of the others. Fodor claims that teleological accounts of meaning offer a solution of this form. According to such accounts, there are mechanisms the function of which is to mediate the causal relations between environmental states on the one hand, and mental states on the other (p. 17). These mechanisms do not always function, because sometimes the conditions are not favorable for their functioning (e.g. it is difficult to see the color of an object in a dim light). But the mechanisms nevertheless proliferate because in *Normal* situations they do function properly. Thus in Normal situations 'X's are only caused by Xs, whereas in other situations this need not be the case. Presumably, what a 'Normal' situation is can be specified in naturalistic terms (e.g. the light conditions are such and such, etc.).

We see here how the solution fits the general form: Normal situations are of type 1, abNormal situations are of type 2. Given that the characterization of Normal situations was done in naturalistic terms, we get a naturalistic account of meaning that seems to avoid the disjunction problem: 'X' means X if

1. Xs causes 'X's.
2. In *Normal* situations, only Xs causes 'X's[5].

Let us look at an example to clarify the above: a frog snaps its tongue at a fly if there is a state of its nervous system that is caused only by flies when the circumstances are Normal. This state thus means *fly*, and it means *fly* even when the conditions are abNormal and it is actually snapping, mistakenly, at something else.

Fodor claims that this only seems to solve the disjunction problem; but it doesn't, really. Suppose that we claim that the frog is not really snapping at

[5]The careful reader will probably have noticed that this formulation of the conditions set by teleological accounts is different from the formulation I had offered of Millikan's conditions at the beginning of this chapter. This difference is crucial, and I will return to it later on.

flies. Instead, it is little black dots that cause the neural state which makes the frog snap. So this state is about little black dots, not about flies. The problem for teleological theories of meaning is that they cannot decide between the two possibilities, on selectional grounds. Since flies are in fact black dots, a Normal situation in which only flies cause snaps is *ipso facto* a situation in which only black dots cause snaps. A frog that will snap at black dots will thus end up eating the same number of flies as a frog that snaps at flies, and will survive just the same. 'Darwin doesn't care which of these ways you tell the teleological story' (p. 20). You cannot distinguish by appeal to evolutionary considerations, Fodor claims, between the two kinds frogs, whereas you do need to distinguish between the two kinds of intentional states.

The way to solve the disjunction problem, Fodor thinks, is by appeal to the asymmetric dependency condition, instead of the Normalcy condition. For the example we are considering, this will decide between the two possibilities: the frog is snapping at black dots, not at flies. Why is that? Because 'the frog's snaps at flies are asymmetrically dependent on its snaps at little black dots' (p. 51). That this is the case is established by the observation that frogs continue to snap at black dots that are not flies, e.g. bee-bee pellets, even after 'they have plenty of evidence that the bee-bees that they're snapping at aren't flies' (p. 51). We have asymmetric dependence because frogs will continue snapping at black dots even if there are no flies around, but only bee-bees. Yet they will not snap at flies, if they are not observable as black dots.

5.4 Some problems with Fodor's arguments

Let us accept, for the time being, that Fodor is right: that Millikan's theory doesn't solve the disjunction problem, and that the frog is really snapping at black dots, according to the asymmetric dependence condition. This raises some

worries. First, let us look back at the analogy between intentional and biological categories. Aren't the biological categories having a similar problem to that of the intentional ones? For example, consider the heart's function. Is its function to pump blood, or to pump liquids? We cannot really tell, because in Normal conditions, the liquid that hearts pump is, in fact, blood. And the human body doesn't really care, for survival purposes, under which description it is getting its blood supply. In Normal conditions (where what the heart pumps is blood) the same number of people will survive either way. So, biology also has a disjunction problem. This, admittedly, doesn't show that Fodor is wrong about anything—it may be that Millikan's description is just not good—not for intentional categories, nor for biological ones. But I think it is good, at least for biology.

Second, if Fodor is right in his conclusion that frogs are snapping at black dots rather than at flies, then something is wrong with our intuition (well, at least my intuition) that the frog is mistaken in some sense when it is snapping at bee-bees. But Fodor thinks that this result, rather than being disturbing, is a desirable one. It is desirable because to attribute a mistake to an organism, he says, it must be the case that this organism can recover from the mistake it is making. As he puts it, in order to say that a symbol token is misapplied to X and correctly applied to Y 'there *must* be worlds, consonant with the organism's psychology in *this* world, in which S tokens are applied to Ys but withheld from Xs' (p. 53). In our example: the frog cannot be attributed with the concept of a fly, because it will always misapply it to other black dots which are not flies. Now it is not clear what kind of recovery is demanded. Suppose the frog can never *see* a difference between bee-bees and flies, but spits out the bee-bees and swallows the flies because it can taste the difference. Compare this with a situation where a person is exposed to the Muller-Lyre illusion: he can never *see* that the lines are not of the same length, but he may find that out by measuring them. Or think of the fact that people perceive the sun as going

around the earth, and discover otherwise only by being told that it is them who are moving, not the sun. Are these not mistakes, and aren't measuring and being told ways of recovering from mistakes? Not according to Fodor, as far as I can see, and this is problematic for his theory. These will not be considered mistakes because in such cases there will never be a recovery 'consonant with the organism's psychology' where the belief will be caused by the object of the belief. Accordingly, tasting, measuring, or being told, cannot be considered ways of recovering from an erroneous belief. These are all alternative causes of the belief, and must depend on its being caused by what it is about; but this, in such cases, can never happen.

We have assumed that Fodor was right in claiming that the asymmetric dependence condition leads to the conclusion that frogs snap at black dots rather than flies, and so far have only evaluated the consequences. Let us question this assumption now, by considering the following alternative: frogs have proliferated, among other reasons, because they managed to catch flies. Their snaps at other small, black objects are therefore asymmetrically dependent on the fact that enough of their snaps were at flies, because otherwise they wouldn't have snapped at anything. Instead, they would have died-out, or would have at least stopped snapping out their tongues and would have developed other, more efficient means for catching flies. Now it may be true that their way of catching flies is by snapping at black dots, i.e. the 'mechanics' of their fly hunting can be described in terms of black dots. But it is the fact that they caught flies, rather than other objects, that enabled their proliferation. It may be true again that this specific frog doesn't 'know', so to speak, that it is snapping at flies, and that it will go on snapping at black dots even if none of them are flies, until it dies. But the frog species would become extinct if frogs were catching bee-bees rather than flies. In 'possible world' talk, in a world in which frogs continued to snap at flies (however they managed to do that), but not at other small, black objects, they would proliferate (both the frogs and the snaps). But in a world

where frogs snapped only at non-flies, the frogs or the snaps would disappear. So snaps at black dots that are non-flies, we see, are asymmetrically dependent on snaps at flies.

When Fodor argues for his asymmetric dependence condition, he says the following: 'maybe all that matters [for semantics] is the patterns of causal dependencies that the pursuit of the [linguistic] policies give rise to. That one kind of causal relation between 'slab's and slabs should depend asymmetrically upon another kind of causal relation between 'slab's and slabs might be enough to explain the robustness of 'slab' tokenings, *however the relations are sustained*' (p. 45, my italics). And again, when pointing out what Skinner was right about, Fodor says: 'semantics depend on a 'functional relation'—a relation of nomic dependence—between symbols and their denotata. *How this relation is mediated*—e.g. that it is neurologically mediated, or for that matter psychologically mediated—*isn't part of the semantic theory*' (p. 45, my italics). This is a good point, and it can be applied to our discussion above: how it is that the frog catches flies, i.e. if it does so by snapping at black dots, at winged black dots, or at buzzing round dots doesn't matter to the semantic question. What matters for semantics, seen from the evolutionary perspective, is that frogs catch *flies*. For this, it doesn't matter what the frog responds to.

Our original question was why, if Millikan's account meets Fodor's sufficient conditions for meaning in certain cases, does he think that it is wrong. We saw now that his main criticism, that it does not decide on the disjunction problem, doesn't hold: it answers the problem, in the frog example, exactly in his terms of asymmetric dependence. But here the puzzle deepens: how is it that applying the same consideration to this example leads the two approaches to opposite conclusions? Why does Fodor's asymmetric dependence condition lead him to the conclusion that the frog snaps at black dots, while Millikan's theory, again for reasons of asymmetric dependence, leads to the conclusion that the frog

snaps at flies? It is time we tried to answer these questions.

5.5 How does Fodor differ from Millikan?

There are, basically, two reasons why Fodor's criticism of Millikan is wrong. The first reason is that what *he* thinks the conditions for meaning are in all teleological theories is different from what Millikan's conditions are. His criticism of these conditions, therefore, does not apply to her account. The second reason, or at least where to look for it, is revealed by the discussion in the last few paragraphs. I have pointed out there that if we consider the asymmetric dependence condition only, we may get conflicting conclusions about what an intentional state means, or is about. This suggests that the two accounts, in deciding the case differently for the example above, have relied upon more than just that condition. We are to look, then, for a difference between the accounts that is not made explicit by the conditions for meaning as we've formulated them above. The significant difference, as we shall see, is not in the satisfaction of the asymmetric dependence condition, as Fodor would claim; Millikan's conditions, I still insist, satisfy this condition. The difference is, instead, in which elements are taken to be meaningful symbols.

The first problem with Fodor's criticism is that he doesn't fully appreciate the significance of the notion of a Normal explanation in Millikan's theory. As he formulated the conditions required by teleological theories, the role of characterizing Normal conditions is to identify 'type 1 situations', i.e. such situations where only X cause 'X's. The problem with his formulation is that it remains "backward looking", so to speak, in considering only the causes for a tokening of a symbol; in that, it misses the "forward looking" element that is the heart of Millikan's teleological theory. It misses the consideration of what a device is *for*, i.e. what its proper function is. According to Millikan, Normal conditions

are, before anything else, conditions that hold *when a device performs its proper function.*

In our example, the proper function of the frog's tongue snap is to catch flies. In the context of the proper function's specification, 'flies' cannot be replaced by 'black dots', while preserving the truth value of the specification: tongue snaps do *not* have the proper function of catching black dots. This is clear because the proliferation of frogs can be explained by their catching flies, but cannot be explained by their catching black dots; frogs that catch only black bee-bee pellets, for example, will not proliferate. We may say that flies are *evolutionarily relevant* to frogs in a way that black dots are not.

But why can't 'black dots' replace 'flies', if flies and black dots are co-extensional when Normal conditions prevail? Isn't Fodor justified in his claim that in Normal conditions a frog that caught a black dot would survive just the same as a frog that caught a fly, because black dots *are* flies? No, the claim isn't justified. The proper function of the snap isn't to catch flies *when the conditions are Normal* (and thus when black dots and flies are co-extensional). The proper function of the snap is to catch flies, always. Even when it is snapping at a bee-bee pellet, the snap's proper function is to catch flies. And since black dots and flies are not *always* co-extensional, but only in Normal conditions, 'black dots' cannot replace 'flies' in the context of the specification of the tongue snap's proper function.

Since it makes a difference what a frog catches, this also defines the proper function of the frog's neural state, S, that causes the snapping: it is supposed to *represent* flies. Which means: it is supposed to enable other devices (such as the tongue snap) that interpret it to perform *their* proper function. In the technical terms used, it is supposed to adapt those other devices to certain conditions so that they can perform their relational proper functions. A Normal explanation

of how S performs its proper function *must* mention the fact that S mapped onto a fly. That, because it is to adapt the tongue snap to conditions such that there are *flies* in the vicinity. Even if S wasn't caused by a fly, but a fly was there anyway, S could perform its proper function. It cannot do so if there is only a non-fly black dot in the vicinity. It is true that a detailed Normal explanation will also mention the fact that S was caused by a black dot; and that it was caused by an excitation of the frog's retina; and by an excitation of the optic nerve, etc. Normal *causes*, true, are too plenty to mark out a device's meaning. But what is relevant to the device's performance, and thus to its proliferation, is its mapping onto flies, rather than any other of its causal antecedents.

The latter point about what a symbol-device represents is brought out clearly if we attend to the notion of a *proximate* Normal explanation. As was explained in Chapter 4, Normal explanations can be of different levels of proximity, or detail. The most proximate Normal explanation of a device's functioning only explains how, given its constitution and the prevailing conditions, a device performed its proper function in most cases. It does not reach into the past, to explain how those conditions came about or what causal relations they have with the functioning device. In our example, a proximate Normal explanation of how S represents a fly will only mention that S co-occurs with flies; it need not mention that S was caused by a fly, or by anything else for that matter. Only more detailed Normal explanations will also mention the causes for S's production. The proximate Normal explanation thus picks out, from several properties common to a fly, the unique property which is the meaning of S. Only that property, of being a fly (rather than, say, a black dot), will have to be mentioned in the proximate Normal explanation as what S must map onto. That it also incidentally maps onto black dots is not essential for S's performance, and will not be mentioned in the proximate Normal explanation. We may thus define the meaning of a symbol as what it should map onto according to the most proximate Normal explanation of a device's proper functioning (as in fact

Millikan defines it, see (Millikan 1984, p. 100). In my formulation of Millikan's first condition, this is captured by the requirement that 'X' *must* Normally map onto X.

The specification of the *causes* for a frog's tongue snap, in contrast with the specification of its proper function, does not create a context that discriminates in this way between flies and black dots. It *is* always true that when a fly causes a snap, a black dot has also caused a snap. It is therefore true that *Fodor's* formulation of the conditions for meaning that he thinks all teleological theories set do not solve the disjunction problem. In Normal situations where only flies cause the neural state S (or cause the frog to snap its tongue), it is also true that black dots also cause the same thing. But Millikan's formulation is different, and does answer this problem.

Fodor, however, argues also against the idea that the *function* of a device can be uniquely identified, not only against the idea that the causes that enable the performance of a given function can be distinguished. He says that snapping at flies rather than black dots would have a selectional advantage only in a counterfactual world where there wasn't a reliable connection between the two properties. But, he says, 'it is, to put it mildly, not clear how utility that accrues only in *counterfactual* environments could produce selectional advantages...'. 'What selection wants is that some actual frogs should actually go hungry in consequence of actually snapping at the wrong sort of things' (pp. 23–24). In other words, Fodor is saying that an evolutionarily based theory cannot appeal to counterfactual selections, only to real ones. I'm not sure that he's right about that, but even if he is, he himself offers a way out of this problem. He argues that it is methodologically 'a bad idea to require of philosophical analyses that are articulated in terms of nomic relations among properties that they be, as one says in the trade, 'cashed' by analyses that are articulated in terms of counterfactual relations among individuals' (p. 40). For the notion of

asymmetric dependency, for example, it is enough that 'we can know that there are asymmetric dependencies among nomic relations between properties without knowing much about which counterfactuals these asymmetric dependencies make true' (pp. 41–42). This holds similarly to the notion of evolutionary relevance. It is enough that there is a nomic relation between frogs and flies, and not between frogs and black dots, i.e. that frogs need to eat flies to survive, but don't need to eat black dots, to make sense of the claim that flies are evolutionarily relevant to frogs in a way that black dots aren't. We do not need to translate this claim into counterfactuals. You don't need, thus, a 'counterfactual frog going around hungry' to make flies, rather than black dots, relevant to frogs' survival. You don't need this any more than Fodor would need an actual asymmetrically dependent misapplication of an 'X' token, for 'X' to mean X.

Fodor has to accept this point about evolutionary relevance, if only because it can be given in his own terms of asymmetric dependence: a frog's snap at black dots which aren't flies asymmetrically depends on its snap at flies, as I've explained earlier. Let me stress again that this relation of evolutionary relevance between frogs and flies holds whether frogs 'know' it or not (i.e. whether frogs can ever discriminate between flies and black dots), and whether we know it or not. Evolutionary relevance is a relation at least as real as the asymmetric dependence is, if only for the reason that the first can be explicated in terms of the latter. It is, however, a more restricting condition than just the requirement for asymmetric dependence. It demands that the asymmetric dependence will hold because of evolutionary considerations.

Let us turn now to the second problem with Fodor's criticism of Millikan. I have said above that the fact that Fodor and Millikan decide differently on the content of the frog's mental state in the above example indicates that the apparent agreement between their accounts must be superficial, concealing some controversial underlying assumptions. I will now try to reveal the difference

between the accounts, i.e. what additional implicit assumptions are being made. When Fodor discusses the notion of a recovery from a mistake (see p. 130 above), he says that for a symbol 'X' to mean X there must be circumstances, *consonant with the organism's psychology*, in which he will apply the symbol to X, but will not apply it to Y. The particular organism's psychology, then, is taken by Fodor to be constitutive of the meaning of its mental states. Millikan's account, instead, takes into consideration the survival conditions for the species over time in defining the content of its intentional states, and ignores the psychology of the particular organism[6].

An important point revealed by observing this latter difference, is that Fodor is thinking of meaning, or content, as something that can be defined over an individual organism on the basis of *its* causal relations with other parts of the world only. According to Fodor, to find out what is represented in *this* frog's brain we experiment with it by shooting bee-bees around it and checking how it responds. The role of evolution theory (and Fodor doesn't deny that it has a role, only not in semantics) is to explain how the frog evolved so that it has a certain mechanism which has such and such discriminatory powers. But it is the actual discriminatory powers that we're after when we're looking for the content of this individual brain, and we can put evolution aside.

In contrast, Millikan's theory of content emphasizes the role of evolution theory in the very definition of the content of an intentional device, not just in explaining how something that has this and that content came into being. As Millikan puts it, the proper function of a mental intentional state (among

[6]It should be noted that obviously not all intentional states are like the frog's fly detectors, in being "wired in" and designed by long evolutionary processes. Instead, the meaning of mental concepts that people learn is defined by their *derived* proper functions. These are derived from the devices that have the proper function of forming concepts, devices that were, supposedly, designed through evolution. More on that can be found in (Millikan 1984) and (Millikan 1989). Public language terms, again, have a different sort of history behind their proper functions. For our discussion here, however, these differences in the source of the proper functions, and whether they are direct or derived, is not relevant. What matters is only that meaningful devices have proper functions that determine their meanings.

other things with proper functions) 'has to do not with its powers but with its history' (Millikan 1984, p. 17). Thus, an experiment that puts one frog in an environment with only bee-bees around it shows nothing about its mental contents, however the frog responds. Instead, if you experiment with a large population of frogs over a long enough time, Fodor's counterfactual claims do not hold: it is not the case that 'frogs continue to snap in worlds where there are dots but no flies'. *This* frog may do so, for a while (until it dies); but not Frogs. And if it is the case that 'they don't snap in worlds where there are flies but no dots' it is for the trivial reason that they will not exist in such a world. What this experiment might reveal, though, is that they *will* snap in a world where there are flies but not black dots. This might be the case because they will have evolved a new mechanism that enabled them to *continue* to detect flies, not because then they would have a different mental state.

For Fodor, the meaning of a certain symbol token 'X' (e.g. a certain mental state) is determined by *its* intrinsic nomic relations with properties that may bring it about, given that it is embedded in a certain sign-producing mechanism (the organism's 'psychology'). To mean X, 'X' must be such that it can be caused by X, or if it is caused by anything else, this must depend on the fact that it can also be caused by X. 'X' shares with *all* tokens of its type this kind of nomic relations with the relevant properties that may cause it. For Millikan, instead, a symbol 'X' is classified as one not because of any intrinsic nomic relation *it* has with such properties, but because it belongs to a certain reproductively established family. Each member of that reproductively established family, in turn, means X only because enough of the members (though not necessarily all) actually had certain causal relations with X, relations that were a Normal condition for the performance of their proper function.

Notice that there are two differences involved here between the accounts. First, a difference in whether or not all tokens of a symbol must have the relevant

nomic causal relations (or mapping relations) with the properties they denote, and with other properties that may cause them. Second, the fact that Fodor is only demanding *nomic* relations, rather than actual ones, between a symbol and its causes, whereas Millikan demands actual cases where these causal (or mapping) relations have been instantiated. This difference is pronounced, for example, by Fodor's theory allowing for the possibility that there will only be one token of a certain symbol type, i.e. that a symbol will mean something only once (this raises an interesting question that I will not go into, of whether a symbol can be used only once, and be misapplied on that occasion). In Millikan's theory, this is not possible.

To make these crucial differences between the accounts explicit, a change must be made to the first condition of both. Fodor's first condition should read

1. *For all tokens of* 'X', 'X' *can be* caused by X.

whereas Millikan's first condition should read

1. *For some tokens of 'X'*, 'X' *did* Normally map onto X[7].

To summarize: When Millikan says 'a symbol type 'X' means X' she refers to a type which is a reproductively established family; tokens of this type do not necessarily share the same nomic causal relations with different properties. In other words, a property that causes one X token of that type need not necessarily cause other X tokens. In contrast, when Fodor talks of 'a symbol of type 'X' ', the type includes only symbol tokens which are embedded in identical, or at least functionally equivalent symbol-producing mechanisms, such that the same nomic

[7]These additions to Millikan's first condition are actually superfluous. The basic formulation, 'Tokens of 'X' must Normally map onto Xs', already allows that not all tokens of X did so map, and assumes that some tokens did actually map onto X in the past. The addition is made only to emphasize the difference between Millikan's and Fodor's formulations.

causal relations are maintained for all tokens of that type. These differences between the accounts makes Millikan's definition of meaning apply to a different class of symbol tokens than that which Fodor's account applies to.

It is these crucial differences that account for the divergence in their analyses of the frog-fly-black-dot example. We see that Fodor's criticism of Millikan, that her account doesn't answer the disjunction problem, depends on this implicit assumption that he makes of what counts as a symbol token of a certain type. His claim, for example, that frogs cannot recover from the habit of snapping at bee-bees, is true only under this assumption. The dispute between them is over what the definition of meaning should apply to. Many considerations can bear on the problem of which account should be preferred, but here I will only discuss one, the question that Fodor himself has raised: which of the accounts manages to solve the disjunction problem?

5.6 Who solves the disjunction problem?

I would like to argue that Fodor's asymmetric dependence condition, that was designed to solve the disjunction problem, fails to do that in fact. Let us have a closer look at Fodor's two conditions of meaning, with his example of dogs, and of cats-on-a-dark-night. Fodor claims that if dogs cause 'dog' tokens and so do cats-on-a-dark-night, the fact that the cats-on-a-dark-night causing 'dog' tokens are asymmetrically dependent on dogs causing 'dog' tokens makes *dog* the meaning of 'dog'. But the fact that a dog that I encounter now (but never before) causes a 'dog' token surely depends, asymmetrically, on the fact that other dogs have, in the past, caused 'dog' tokens. This opens the possibility that 'dog' means *dogs-that-have-caused-'dog'-tokens-in-the-past*: the dog I encountered now doesn't belong to that category, but it has caused a 'dog' token, and in a way that was asymmetrically dependent on the tokenings of 'dog' caused by dogs-that-

have-caused-'dog'-tokens-in-the-past. In fact, another category that 'dog' may mean is even narrower: it is dogs-that-have-caused-ME-'dog'-tokens-in-the-past. Again, my present tokening of 'dog' asymmetrically depends on tokenings of that category. Fodor's asymmetric dependence condition, we see, doesn't decide between these different properties, and thus doesn't solve this version of the disjunction problem. This, by the way, is not just a technical objection based on circumstances that cannot really ever happen. It can happen quite easily, for example when a person is introduced to a new class of objects and mistakenly takes the few exemplars shown to him to be the entire extension of that class. When he sees another, new object from that class he might well mistake it for one of the objects he already saw and for *that* reason apply the same concept to it.

One possible reply to my argument may be the following. Fodor added a caveat to his above conditions for meaning, saying they must apply with 'a synchronic force'. That is, the asymmetric dependence in the definition should be understood as referring to a *current disposition* to apply 'X' to Y's that depends on a *current disposition* to apply 'X' to X's[8]. This caveat may seem to block my objection, because I'm referring to dogs that have caused 'dog' tokenings in the past. But in fact this does not affect my argument: the reference to the past is part of the property, not the dependency. In our example, my disposition to apply 'dog' to this new dog does depend *now* on my current disposition to apply 'dog' to dogs-that-have-caused-'dog'-tokens-in-the-past.

Another reply to my argument was made by Fodor, in a personal communication. Since the connection between having the property X and causing the tokening of an 'X' is supposed to be a *nomic* connection, he says, this rules out X being any property the possession of which *entails* being a cause of 'X' tokens. Thus it rules out X being the property of being a cause of 'X' tokens.

[8]See (Fodor 1987). Why Fodor needs this caveat is explained on p. 109, but it is irrelevant to our discussion here.

This claim, however, does not answer my objection. The property I have suggested was 'dogs-that-have-caused-'dog'-tokens-*in-the-past*'. The possession of this property does not entail that a dog that has it *is* a cause for 'dog' tokenings now. The connection, therefore, between this property and the property of being a cause of a 'dog' token (if there is one) is a nomic connection, not a logically necessary one, as required.

Does Millikan's theory run into the same problem? I think not. The first condition set by her theory picks out, from the many properties common to the class of things that may cause the tokening of a symbol, the property that is relevant to its proper functioning. Thus, in the above example of the tribe of cow hunters, it is the fact that 'cow' co-occurs with cows, not necessarily with any specific cows (e.g. ones encountered before), that figures in an explanation of the functioning of 'cow' tokens. That cows which figure in a Normal explanation for the functioning of 'cow' tokens could also be described as cows-that-caused-'cow'-tokens-in-the-past is just as irrelevant to the Normal explanation as the fact that flies could also be described as black dots.

We must conclude, then, that it is Millikan's account rather than Fodor's that solves the disjunction problem. Different considerations may give some points to Fodor's views, some to Millikan's. But since Fodor has put the disjunction problem as the decider between naturalistic theories of meaning, Millikan has won a game for which Fodor had provided the ground rules.

Chapter 6

Millikan's Account of Metaphor

6.1 Introduction

In the following two chapters I will tie together the two topics I have discussed in the former chapters: metaphor, and Millikan's evolutionary theory of meaning. That is, I will discuss the problem of metaphorical meaning, or more generally, the question of systematicity in metaphorical interpretation, in terms of Millikan's theory of meaning. Before I begin, I should explain the motivation for making this connection: why do I expect Millikan's theory to provide any insight into metaphor? At the beginning of Chapter 4 I mentioned two reasons for applying Millikan's theory to the question of metaphorical meaning: the clear distinction it makes between speaker's intentions in communication, and the meaning of language devices in themselves; and its focusing on the history of language devices in determining their meaning which is relevant for clarifying the notion of a fresh metaphor. Two further reasons are the following: The first has to do with the notion of regularity that has been my main concern in discussing metaphor, reflected in the puzzle I've been calling 'the basic dilemma'. The puzzle, recall, was whether there is a systematic relation between what we would intuitively call a fresh metaphor's metaphorical meaning, and its literal

meaning.

Millikan's notion of proper function brings out the difference between general, statistical regularities about language devices, in contrast with regularities that hold only in particular cases, where these cases can be independently distinguished as relevant. The distinction is made by identifying the evolutionary relevant, proper function of a device, and discriminating between the device tokens that perform that function and those that don't. Semantic claims about regular correspondence relations that hold between symbols and their denotata, for example, need not be true of all, or even most actual symbols that are produced. They should hold only in the case of symbols that are successfully performing their function, where this function depends on such correspondence relations. Accordingly, a semantic theory cannot start off by just describing observed regularities between symbols and things, but must hypothesize about the functional role of the symbol-devices it is describing. From this perspective, it is hoped that discussions about the meaning of metaphor can make progress by not just arguing over whether, as a matter of fact, there are statistically valid facts about regular relations between expressions and their metaphorical meanings. If the proper function of metaphor can be identified, a sub-class of metaphors can perhaps be identified of metaphors that do perform their function. It will make more sense to discuss whether in *these* cases there are any regular relations between expressions and their metaphorical meanings. This line of argumentation will be explored in the following chapter. In this one we shall see other ways in which the notion of proper function may play a role in the discussion of metaphor.

The second reason for expecting Millikan's theory of meaning to provide insight about metaphor is more general, and related to the second reason I've discussed in Chapter 4. Her proposal is based on the theory of evolution, and thus considers meaning from a historical, diachronic perspective. Moreover,

evolution theory has to do not only with the emergence of stability in an unpurposeful world, but also with change and variety. Metaphor seems to underlie many diachronic changes in meaning—this is probably one reason that studies of metaphor find it difficult to dismiss it as a phenomenon that is not relevant to semantics. An account of meaning that is based on a historical perspective, and in particular on evolution theory that is concerned with diachronic change, seems therefore not only a promising direction for inquiry about metaphor, but also one that provides a principled motivation for taking a diachronic approach to semantics. I will now follow, however, the first kind of motivation and consider some ways in which Millikan's notion of proper function is relevant to metaphor.

6.2 Metaphorical meaning as derived proper function

Millikan, it should be stated from the outset, is not really concerned with giving a detailed account of metaphors and their meaning. I will reconstruct, however, such an account on the basis of the sporadic remarks she makes throughout her book (Millikan 1984) on the subject[1]. Millikan's views will be presented mostly in comparison with Searle's account of metaphor, because as will become apparent, this reconstructed account closely resembles Searle's. But, I will argue, it manages to avoid some of the difficulties that I have pointed out in the latter.

Millikan's first remark on metaphor is the following: 'language devices are often used in secondary or parasitic ways, as in metaphor, sarcasm, or lying' (Millikan 1984, pp. 4–5). This, more than anything, should be seen as an "ideological" statement: Millikan is taking sides on the issue of whether metaphor is

[1]For ease of exposition, I will refer to this reconstructed account as Millikan's account of metaphor. Because of the large amount of my interpretation that is involved, I want to stress that it may not faithfully reflect Millikan's own ideas.

or isn't a *semantic* phenomenon, and places it outside semantics. It is instructive to note here both the similarity and the difference of her claim to Davidson's views. They both agree that metaphor is a case of deviant use, and that it does not involve a change in meaning. But there is an important difference in what they mean by these claims: for Davidson, it is enough that metaphor is a matter of *use* rather than meaning (in the sense of correspondence to certain truth conditions) to make it irrelevant as a semantic phenomenon. The fact that it is a *deviant* use, is subsidiary to the claim about semantics (as I've explained in Chapter 2). For Millikan, instead, use and meaning (in the sense of correspondence to aspects of the world) are entwined together—proper function, a linguistic device's use in a sense, *is* one aspect of meaning. She sees metaphor, therefore, as a matter of *deviant* use. It is the stress on deviance, rather than use, that licenses a semantic account not to apply to metaphors.

As was explained above, in chapter 4, tokens of language devices have dual proper functions. As members of reproductively established families, they have a *direct* proper function; as items produced by purposeful speakers, they have a *derived* proper function—derived from the speaker's intentions and purposes. We can learn more about Millikan's view on metaphor from what she says about these dual functions: 'The direct and the derived proper functions of a language [device] token often are not the same, for example, whenever a speaker does not understand how to use a word or uses it sarcastically or metaphorically' (Millikan 1984, p. 49). It is here that Millikan's account resembles Searle's. She is saying that a metaphorical use is one where the speaker's intended meaning diverges from the sentence's own meaning; and apparently what we treat as metaphorical meaning is the speaker's intended meaning. However, it should be stressed that Millikan's idea of meaning is quite different from Searle's; the consequences of this difference will be discussed later on.

Let me expand on what happens to the direct proper functions of language

devices (their 'own' meaning) when used metaphorically. First, since these functions are determined by the devices' history, rather than the way they are being used on a particular occasion, they remain unchanged. Second, an expression used metaphorically will, most probably, fail to perform its direct proper function. In the case of an indicative sentence used metaphorically, for example, it will not create in the hearer the true belief that it should (given the proper function of the words that comprise that sentence), for at least the reason that metaphors are usually false. But even in the case of true indicative sentences used metaphorically, e.g. 'no man is an island', a metaphor cannot perform its direct proper function Normally. That is, even if the hearer came to the true belief that no man is an island because of hearing this sentence, the sentence did not perform its proper function Normally, because a Normal explanation would require that the sentence's derived proper function (determined by its utterer's intentions) must coincide with its direct proper function. Since the speaker did not purpose that the hearer should come to this belief, the derived and direct proper functions did not coincide, and thus the sentence couldn't have performed Normally. This would mean that even though it did perform its proper function, this wasn't one of the stabilizing cases where the fact that it performed its proper function would explain why this use was repeated. More intuitively: since the speaker didn't intend the sentence to function as it should, the fact that it did so function wouldn't encourage him to continue to use it for that specific purpose.

Discussing a metaphor's direct proper function calls for explaining the notion of literal meaning. An expression's direct proper function and its literal meaning, it should be noted, are not the same thing. Technically, according to Millikan, there is no such thing as literal meaning (as opposed to non-literal meaning), since language devices have only one kind of meaning, i.e. their direct, stabilizing proper function. There is just literal *use*, and even this use does not necessarily correspond to cases where a linguistic device actually performs its stabilizing

proper function. For example, when one is lying he does not purpose that the indicative sentence he is using should perform its direct proper function of creating a *true* belief in the hearer. Nevertheless, he does purpose that his sentence will be *taken* literally, i.e. that the hearer will respond with a reaction that will have the same proper function as that which the sentence has. Put more intuitively: using a sentence to lie isn't a case of using it non-literally. A non-literal use is only one where the hearer isn't intended to take it literally.

Metaphorical uses of sentences, Millikan says, are parasitic uses. That a sentence used metaphorically performs its derived proper function depends on a relation it bears to its direct proper function. In some cases, Millikan says, this relation is a conventional one, copied from one generation to the next. She insists, however, that even though the relation between the direct and derived proper functions is conventional, the intended function cannot be attributed to the sentence itself. That is because, taken alone, such parasitic uses as the metaphorical ones cannot sustain the proliferation of the language devices involved. Instead, metaphorical uses depend for the successful performance of their derived proper functions on the language devices that are being used having their own, direct proper function.

Millikan says that sometimes, however, metaphorical uses do not depend on such conventional relations, but are instead improvised uses. She doesn't expand on this claim, and understanding it will need some reconstruction. Millikan does discuss the notion of an improvised sign: it is something that does not have a direct proper function because of not being a copy of other signs. Accordingly, it cannot be an intentional icon (i.e. a standardized sign), but it is nevertheless used in a manner similar to that of intentional icons, to elicit certain responses in a hearer (e.g. making him acquire a belief, or act in a certain way). An example for an improvised sign could be someone's inventive use of hand gestures in communication. Such improvised signs have only derived proper functions—

they mean whatever their producer intends that they mean, and no more. The general, Normal explanation of how one communicates via improvised signs will mention the fact that the producer of the sign has designed it in such a way so that it will fit some general interpretative mechanisms he assumes the interpreter has. It will not, however, depend on any specific mechanisms that have been standardized to interpret a specific kind of sign, as in the case of using standard language devices. For example, if one used a rotary movement of his hand to signal something about a car, he would be depending on his hearer's general interpretative capabilities, rather than on that gesture having a standardized role in communication.

Now, if we consider metaphors again, surely Millikan doesn't mean that there are metaphors that do not have their own, direct proper functions; after all, metaphors are always constructed of words. In that sense, even fresh metaphors must still be cases of parasitic usage of words. In other words, even fresh metaphors have dual proper functions, in contrast with improvised signs. What Millikan must mean is that like improvised signs, some metaphors (the fresh ones) do not have conventional ways of being used. Instead, speakers rely on some general, non-specific mechanisms that a hearer is assumed to have, over and above their reliance on those mechanisms that have been standardized. This is again similar to a claim Searle has made, that since 'the knowledge that enables people to use and understand metaphorical utterances goes beyond their knowledge of the literal meanings of words and sentences, the principles we seek are not included ...within a theory of semantic competence as traditionally conceived' (Searle 1979, p. 78).

To summarize Millikan's position on metaphor: a linguistic device , in itself, cannot be classified as a metaphor. Instead, a metaphor is a linguistic device token that is intended by its utterer to perform a function that is different from its own, stabilizing proper function. The function intended by the speaker is

the linguistic device token's derived proper function, and corresponds to what we intuitively take to be an expression's metaphorical meaning. A single, deviant use of a linguistic device does not affect its stabilizing proper function. Metaphors are parasitic uses of language devices, in the sense that their performance of the derived proper function they have is dependent on their having a direct proper function, but their performance of the derived proper function doesn't alone sustain their proliferation. The performance of a metaphor's derived proper function may depend on a conventional relation it bears to its direct proper function, or may depend instead on general interpretative abilities of the hearer, which haven't been standardized.

6.3 Examining Millikan's account of metaphor

Millikan's account of metaphor, as presented above, may not strike one as excitingly different or innovative, compared with the treatments we've encountered above. It seems that if you replace 'direct proper function' with 'sentence meaning', and 'derived proper function' with 'speaker's meaning', you get Searle's account. But this apparently small difference makes all the difference. It marks the distinction between an expression's *being* understood in a certain way, and its being *supposed* to be understood in a certain way, which is no minor difference.

In the following, I will evaluate Millikan's position, mostly by checking it against criticisms directed at Searle's apparently similar account. I intend to show that their different understandings of the notion of meaning enables Millikan to withstand most of these criticisms. In particular, it allows her to *consistently* claim that metaphorical meaning is defined by the speaker's intentions. But first, let me consider a difference between their treatments that might elicit objections to Millikan's view.

6.3.1 Conventional metaphors

As I have mentioned above, Millikan (unlike Searle) allows that conventional metaphors would be treated as genuine metaphors. In her terms, it means that in at least some cases of using conventional metaphors, their derived proper function will still differ from their direct proper function. The basis for this claim is that she thinks that what is conventional about them is not their metaphorical meaning, but the way their standardized proper function is used to perform a different, though related function. In other words, she thinks that the language devices involved are copied not because of their having an alternative standardized proper function, but because their standard proper function is found suitable for performing some other, derived function.

This position could be objected to for the following reason: if a regular relation is sustained by convention between an expression and its secondary, derived effect, why is such a two-stage, "dog-legged" explanation needed? Why not assume that the metaphor just means what is conventionally calculated to be its meaning? This is one place where Millikan's notion of meaning as direct proper function, rather than an established regularity, plays its role. A device's direct proper function, recall, has to do with the reason for its proliferation, not just with its regular performance. Now consider this analogy: suppose that screwdrivers, apart from being used to manipulate screws, are used quite regularly to poke holes. Suppose also that this alternative use is being copied rather than invented over and over again. The screwdrivers, however, retain their distinctive shape because this shape is suitable for manipulating screws, not because it is particularly suitable for poking holes. Presumably, if poking holes was the reason for their being copied, their shape would have changed accordingly to be more suitable for that purpose, and they would be pointed rather than flat headed. The fact that they did not so change, would indicate that the screwdrivers' being used quite regularly to poke holes does not bestow

them with the direct proper function of poking holes. This use, so to speak, is a parasitic use they have been put to; it is parasitic upon the shape they have as *screwdrivers*, but poking holes, alone, would not sustain their having that particular shape copied.

Similarly, language devices may be used, conventionally, for achieving certain metaphorical effects. Nevertheless, as long as these effects cannot alone sustain the proliferation of these language devices, the language devices cannot be said to have acquired an alternative proper function. Now, we know that words do acquire, through metaphorical use, alternative meanings. Their being used, over and over again, for some parasitic function, may finally result in their acquiring an alternative proper function. But the fact that they are regularly used for an alternative purpose cannot, on its own, imply that they *have* acquired an alternative proper function, or meaning.

This reply accords, I believe, with Millikan's view[2]. I must, however, express some reservations I have about it. In the analogy with the screwdriver, the claim that poking holes was a parasitic, rather than a proper use of screwdrivers was made plausible by the change in the shape that would follow a change in their proper function. The functioning of words, however, doesn't bear a direct relation to their form. Thus, when an established metaphor does acquire an additional or alternative direct proper function, this is not reflected in a corresponding change in its form. It is impossible to tell, therefore, when a word has actually acquired a new proper function, rather than being used parasitically. Millikan can hold her ground by insisting that this is only a question of *knowing* if a word has, or hasn't acquired a new direct proper function, but that there is a matter of fact whether it has. I am not sure, though, what kind of fact it is that sustains this difference.

[2]In fact, the example of the screwdriver was suggested to me by Millikan in a private communication (though not exactly in this context), and I am grateful to her for this contribution.

Searle, however, who did not distinguish between regular and literal meaning, was forced to exclude any convention governed metaphors from being genuine metaphors. For the same reason, he was prone to the attack that if systematic relations did in fact hold between literal and metaphorical meaning in the case of *fresh* metaphors, as he described in his metaphorical interpretation principles, there were no reasoned grounds for not attributing the meaning to the words, rather than to the speakers; which leads us to consider Millikan's account of fresh metaphors.

6.3.2 Fresh metaphors

My main criticism of Searle, recall, was his inconsistency in claiming that metaphors mean what their speaker intend that they mean, but at the same time trying to impose limits on what a speaker may mean metaphorically. These limitations were needed in order to explain why these intentions are successfully carried out. Millikan avoids this inconsistency by dropping the limitations on the speaker's intentions. A speaker can mean just *anything* by his utterance (any utterance, not just of metaphors), and his intentions or purposes will determine the utterance's derived proper function. Millikan can allow that, because a device token's having a proper function isn't a guarantee for success in performing it. Having a (direct or derived) proper function only determines what a device *should* do, but may fail to do. In particular, a linguistic device token has the derived proper function it does not because it *can* perform what the speaker intended that it should, but rather because enough earlier tokens of language devices that this speaker had produced did manage to bring about what he has intended. For example, the nonsensical utterances of a person that had spoken normally up to a certain stage, but was hurt by an accident that impaired his speech faculties, will still have (at least for some time) the derived proper function determined by his intentions. It is not his intentions that are faulty (as

might be implied by Searle's account), but the speech producing device that has *failed.* Conversely, utterances produced by a parrot, though they retain their direct proper functions, *never* have any derived proper functions because we assume they are not produced with the right sort of purposes. Millikan, we see, manages to make a consistent claim about metaphors' having derived proper functions (and meaning, in this sense) that are determined by their utterer's intentions. This is achieved by her notion of proper function, that replaces the need for giving an account that will explain why most, or even all metaphors work.

Millikan's account doesn't only free her from a commitment to explain *why* most metaphors work, it also doesn't require that she explains *how* metaphors work. I have criticized Searle's attempt to describe, in a uniform way, the psychological processes that underlie the interpretation of metaphors. As I've pointed out, there is no reason to assume any uniformity in these processes, let alone attributing to them any psychological reality. Millikan avoids this kind of problems. As became apparent in her discussion of intensions, she does not even assume that in cases of standard communication, the underlying psychological processes are uniform. What remains constant is an expression's proper function and sense, i.e. its Normal mapping value, rather than the way an individual recognizes what this value is. The case of fresh metaphors, as improvised uses of language devices, is even stronger in this respect: there is no reason—and no need—to suppose that there is a uniform way metaphors perform their derived proper functions. All that is required, say in the case of a sentence used metaphorically to convey information, is that this information *is* conveyed, i.e. that the hearer will acquire the belief he was intended to acquire by the speaker. Even in the case of conventional uses of metaphors, what is copied is the (parasitic) *use*, not the interpretation process. In other words, the associative links that underlie each hearer's correct understanding of a metaphor's intended meaning may vary between hearers.

A more general advantage of Millikan's account over Searle's is the fact that her claims about the divergence, in the case of metaphors, between a sentence's *own* meaning, and its intended meaning, are based on a very detailed account of what it is for a sentence to have its own meaning, and how its having a meaning is distinct from the speaker's intended meaning. Searle, like Millikan, employs the distinction between sentence meaning and speaker's meaning in the main thesis of his paper on metaphor; but unlike Millikan, he does not explain what sustains this distinction. His above quoted claim, therefore, that metaphors involve the use of non-semantic knowledge, is a dogmatic claim, if not downright circular. Obviously, if metaphorical utterances have, in addition to their literal meaning a non-literal meaning, semantic knowledge will not do to interpret them. But *why* is their additional, intended meaning, not part of the semantic theory? This is especially unclear, as I've pointed out, given that Searle thinks the intentions are so well structured.

6.3.3 Cooper's objections

Cooper, in his criticism of Searle's position on metaphor, made what he called 'the perversity objection'. He argues that the use of metaphor 'is such a familiar and ubiquitous ingredient of speech that, if it were [a perverse use], few stretches of everyday conversation would escape the presumption of censure... Consequently an ideal analysis of metaphor should have as a corollary that metaphor is not perverse' (Cooper 1986, p. 78). He then offers a detailed analysis of Searle's account that shows how it cannot escape entailing that metaphor is perverse, in violating one of several 'conversational maxims' (see Grice 1975).

Millikan, as I've pointed out earlier, treats metaphor as being a case of deviant use. Prima facie, she seems therefore prone to this kind of objection: if the utterers of metaphor do not intend that it means what it does mean, too

many cases would be a violation of a conversational maxim to the effect that one should mean what one says. Cooper calls this maxim 'the transparency convention', and argues that it is needed 'since suspension of the assumption that people are generally obeying it would make attempts at interpretation nugatory' (Cooper 1986, p. 81). But Millikan can be seen to reject that principle. In a sense, she is making an even stronger demand than Cooper. Not only is she arguing that, say, indicative sentences should be interpreted in a standard way—she even argues that they should be *true*. Cooper, in contrast, is satisfied with the demand that people should only mean what they say, even if they do not intend that what they say should be taken as true. But Millikan's understanding of the norm she sets, is much weaker than Grice's (and Cooper's) understanding of the norm set by a conversational maxim. The norm set by a device's proper function does not require that *most* cases will comply with it; it only requires that *enough* cases will accord with the norm, enough to "keep the device going". But enough may be much less than a majority.

Let us recall why Grice has brought up the notion of conversational maxims in the first place. In the introduction to 'Logic and Conversation' (Grice 1975) he presents a debate between formalist and informalist approaches to the analysis of the semantics of natural language. The informalist approach argues that there are too many logical imperfections in natural language to allow an adequate logical analysis of its terms. Grice's purpose in that paper is to point out that 'the common assumption of the contestants that the divergences do in fact exist is (broadly speaking) a common mistake' (*ibid*, p. 43). In other words, his strategy for dealing with the embarrassing divergences is to claim that they are only apparent divergences, not real ones.

Millikan's strategy, instead, is to deny that 'too many' divergences are in fact an embarrassment. She rejects the assumption that the statistical norm is the yard stick for evaluating the adequacy of semantical claims, and offers

the notion of proper function as an alternative basis for constructing a semantic theory. Going back to Cooper's transparency convention, Millikan's approach allows her to dispute the need for such a convention. According to her, apparent divergences from that convention need not be undone, to keep the semantical claims intact in most of the cases. They can be accepted as divergences, without harming the enterprise of constructing a semantic theory and without rendering 'attempts at interpretation nugatory'. What I have sketched above is not, of course, a thorough argument for rejecting Cooper's argument, or Grice's ideas. It brings out, however, the significance of the notion of proper function as an alternative to an approach that is perhaps too dependent on statistical norms. In particular, I hope this discussion has demonstrated again the relevance of this notion to the discussion of metaphor.

I have shown shown how Millikan's account could cope with Cooper's 'perversity objection', which Searle was not equipped to answer. I have also claimed that Millikan's account managed to make a *consistent* claim about metaphorical meaning being the speaker's intended meaning, where Searle's account has been inconsistent. The question remains, however, if her account is adequate in other ways, apart from being consistent. Another part of Cooper's criticism of Searle, recall, was the 'indeterminacy objection', where he has said the following:

> '...a metaphor, once announced, belongs like a published poem or an exhibited painting to the world. The speaker, poet or even painter does not have exclusive rights to interpretation...'.
>
> (Cooper 1986, p. 73)

In other words, Cooper is saying that metaphors have a public aspect, and how they are interpreted (over and above their literal interpretation) shouldn't be compared for correctness against their author's intentions. In particular, their being indeterminate in their interpretation isn't a matter of the speaker's intentions: a metaphor may have an indeterminate interpretation even if the

speaker meant something determinate, and vice versa. An account of metaphor that rules out this result, Cooper claims, is inadequate. This criticism holds against any account that attributes to the speaker's intentions what a metaphor means, and Millikan's account fairs no better than Searle's in this respect. In the following chapter, I will therefore discuss the public aspect of metaphorical meaning. I will consider the question of whether there are any standards that determine the meaning of metaphors, that do *not* depend on the intentions of any particular speaker.

Chapter 7

An Evolutionary Perspective on Metaphor

7.1 Introduction

In order to stress her point that a device's proper function depends entirely on its history, Millikan makes the following claim:

> 'Suppose that by some cosmic accident a collection of molecules formerly in random motion were to coalesce to form your exact physical double. Though possibly that being would and would even *have* to be in a state of consciousness exactly like yours, that being would have no ideas, no beliefs, no intentions, no aspirations, no fears, and no hopes (his *non*-intentional states, like being in pain or itching, may of course be another matter.) This because the evolutionary *history* of the being would be wrong. For only in virtue of one's evolutionary history do one's intentional mental states have proper functions, hence does one mean or intend at all, let alone mean anything determinate... That being would also have *no liver, no heart, no eyes, no brain*, etc. This, again, because the history of the being would be wrong. For the categories "heart", "liver", "eye", "brain", and also "idea", "belief", and "intention" are proper function categories, defined in the end by reference to long-term and short-term evolutionary history, not present constitution or disposition'.
>
> (Millikan 1984, p. 93)

In a footnote to this passage, Millikan adds:

> 'A similar situation occurs whenever a mutant appears on the scene. Even though what is new in the mutant may serve a very useful function, hence help the mutant to survive and to produce progeny that survive, etc., still this new aspect has, in the case of the very *first* mutant member of the community, and sometimes for a considerable time thereafter, no proper function at all'.
>
> (*ibid*, p. 338)

My first claim will be that mutants are very different from the imagined cosmic accident; that mutants do have an evolutionary history, and that they do have, as such, a proper function. Accordingly, I will claim that biological devices proliferate not *in spite* of mutations but, to a certain extent, due to mutations. To lend this claim some initial plausibility, let me point out that devices proliferate not only because of being good survivors and replicators, but because of being *better* than others. And devices change and improve (in the strict sense of their potential for proliferation) by mutation and other mechanisms that introduce variance in the replication process; which indicates the relevance of these mechanisms to the proliferation of the replicating devices[1].

In the first part of this chapter, I will focus on this claim as it applies to biological devices, and will identify the proper function of variants. But after all, we are interested here in meaning and, in particular, we are interested in *metaphorical* meaning. So in the second half of the chapter I will return to this topic, and deal with it by extending Millikan's analogy between biological and linguistic devices. My claim is that metaphors are the linguistic analogue of biological variants. Both seem to be 'new appearances on the evolutionary scene', but my claims in the first part of this chapter will serve to examine in what

[1]For ease of exposition, I will use the term 'variant' to encompass all devices that are non-accurate replications of other devices, disregarding the reason for that inaccuracy, i.e. whether it resulted from a chance mutation, or for other possible reasons that will be introduced shortly.

sense metaphors do have, in spite of appearances, an evolutionary history. Accordingly, we will also discover in what sense they have a proper function—and meaning—that is different from the proper function they have as replications. Most importantly, this extended analogy is meant to bring out two things: first, the kind of systematicity that holds between what I shall describe as metaphorical meaning, and the metaphor's original, literal meaning. As it will turn out, this analogy captures the kind of systematicity that should resolve, as I argued in the first part of the thesis, the 'basic dilemma'. That is, it shows how the relation between the metaphorical and original meanings of an expression are constrained, but at the same time open ended. Second, identifying the proper function of metaphor will enable us to restrict our claims about systematicity to those metaphors that do perform their proper function. As in all cases, generalizations concerning a device's functioning cannot be expected to apply to malfunctioning devices.

7.2 Biological mechanisms of variance

7.2.1 The evolutionary history of variants

Millikan's claim that (new aspects of) mutants do not have a proper function, derives from the fact that they are new devices, rather than replications of existing ones. Since 'proper function' is defined in terms of the performance of earlier members of a reproductively established family, mutants as new devices cannot have, by definition, a proper function. I would like to begin by pointing out that the notion of mutant, or more generally the notion of a variant, does not cleanly cut it from its evolutionary history. Saying of a device that it is a variant implies that it has some features that are new, but at the same time relates it to the reproductively established family of devices that it was *supposed* to be a member of. It is related to that reproductively established family of devices

in that a variant can only be defined as a variant *of* some device. A cosmic accident, in contrast, is not related in this way to anything.

Even the notion of a device's having features that are *new*, has implications about the device's aetiology, as the cosmic accident example clearly illustrates. My imagined doppelganger has features that are identical to mine—they are not new in this sense. Its features are new only in the sense of not being a replication of an existing being. I will distinguish, therefore, between a variant *as a new device*, i.e. as a device with particular features that are not replications, and a variant *as such*, i.e. as a device that is different, no matter exactly how, from a specific reproductively established family to which it was supposed to belong, in some sense.

Millikan's claim holds of variants as new devices: no matter how effective the new device is in surviving and replicating, none of its effects can be considered its proper function because it has no earlier replicas[2]. But, as for a variant as such, its relations to its evolutionary history are still apparent. We will need to further explore the nature of those relations, to see if they can serve as a basis for attributing any proper function to variant devices (as such).

I have said that variants are related to some particular reproductively established family because they are supposed to belong to it in some sense, even though they do not share with its members the relevant reproductively established properties. This 'being supposed to' must be explicated again in terms of the proper function of some device or process, in this case the process that controls the replication. In the case of higher order reproductively established families, where the members are not reproductions of one another in the strict

[2]I am making a distinction between the *effects* of a device, and its *function*. Both terms refer to the causal consequences of the device's interaction with its environment, but talking of effects applies to all such consequences, whereas 'function' applies only to effects that are somehow distinguished from the others. They are distinguished, according to Millikan, by the device's evolutionary history; thus, while considering a new device with no history as such, I will talk only of its effects rather than its function.

sense[3], they are members of that family because they are products of a device that has as its proper function to produce other devices with some particular properties in common to all of them. This is the case, as I have illustrated, with mass-produced products: they are similar because they are produced by the same machine, not because they are direct copies of one another. Even if they are somehow different, there is still a clear sense in which they are supposed to be similar, given that they are the product of the same machine with the proper function of producing similar items.

The case of first order reproductively established families is more problematic; its members must be, by definition, reproductions of one another. Which means: they must preserve any change in the model from which they are replicated, unlike the members of a higher order reproductively established family. Xerox copies, we have seen, are reproductions in this sense, and all xerox copies of the same document are therefore members of a first order reproductively established family. Is there a sense in which they are *supposed* to be reproductions of one another, apart from being identical as a matter of fact? Put another way, is there a sense in which there can be a member of a first order reproductively established family that does not share with its members the relevant reproductively established properties? Millikan doesn't discuss this possibility, but I find it important, and of course relevant to the question of whether we can make sense of the notion of a variant of a first order reproductively established family. Only if we can explain how something may be a non-identical reproduction, we can say that something is a variant relative to the model it was supposed to be identical with. This is necessary, because the notion of mutation applies to *genes*, which do form first order reproductively established families. In Millikan's account as it stands, the notion of genetic mutation cannot make any sense—a gene just is, or isn't, a replication of the parent gene. What we need

[3]Recall the distinction made in Chapter 4 between replications and reproductions. Replications may have a common cause which makes them similar, whereas reproductions must be direct copies of the devices to which they are similar.

to do, then, is to somehow relax the demand for identity between reproduced elements.

This, in fact, is quite easy to do given Millikan's theoretical apparatus. Just as in the case of higher order reproductively established families there was a device that had the proper function of producing elements with certain features in common, so can we have a device with the *relational* proper function of producing elements that match other elements, external to the device. This, in fact, is just what we get in the case of the xerox machine: it has such a relational proper function, of producing sheets of paper that match some of the features of whatever is placed onto the machine. Similarly with genes: we may get a variant gene, because there are complex, specialized processes that have the proper function of producing copies of other genes. Even if a gene is not identical with its 'parent' gene, there is still a sense in which, as a product of the reproduction process, the gene is *supposed* to be a copy of it. A variant gene, therefore, is still a member of a specific first order reproductively established family, and has an evolutionary history in this sense, due to the proper function of the reproduction process. It will, for example, have the same direct proper function as its parent gene, even if it cannot perform it due to its varied structure. It should be noted, though, that not all reproductions are carried out by specialized copying processes. Some reproduced elements, e.g. footprints, are just reproductions as a matter of fact. But in such cases, there is no sense in talking of variants, because the reproduced elements are not supposed to be identical; they just are, or are not identical.

7.2.2 Mechanisms of stability

Let us consider what is, exactly, the proper function of the genetic reproduction process, by comparing it with Millikan's example of the pigment arranging device

of the chameleon. They both have relational proper functions: the pigment arranging device to produce a skin color which bears the relation 'same as' to the chameleon's surroundings; the genetic reproduction process, to produce genes which bear the relation 'same as' to the parent genes. Now, the *products* of these devices—a specific skin color, or a specific genetic sequence—have as such *derived* proper functions. Millikan says that the derived proper functions of such products 'are derived from proper functions of the devices that produce them that lie *beyond* the production of these adapted devices themselves' (Millikan 1984, p. 41). Thus, the derived proper function of the specific skin color is to make the chameleon invisible to predators. This is the proper function of the pigment arranging device that lies beyond the production of skin colors that match their surroundings. And the derived proper function of the specific copied gene is to—is to what? What is the proper function of the genetic reproduction process that lies beyond the production of genes that match other genes?

The question is, in other words, how does the organism (and thus the reproduction process within it) benefit from copying the genetic material? Since benefit is measured in terms of proliferation, the answer is in one sense trivial: the copying itself is what enhances the proliferation. But there is a more interesting answer than that. The copying process will preserve in the next generation any adaptive advantages that the parent genes have. Its proper function, therefore, is to produce a gene that will function in the same way as the parent gene has functioned. And the derived proper function of its product, the copied gene, will be to function in the same way as the parent gene has functioned. The reproduced gene's derived proper function is, it turns out, exactly the same as its direct proper function. In this case this seems a superfluous duality, but we will later see that the direct and derived proper functions of reproduced genes do not always match.

7.2.3 Mechanisms of variance

The genetic reproduction processes, as a matter of fact, produce variants. This is due to mutations—"accidents" in the copying processes; but also, in many cases, due to "built in" features of the replication process. In the case of sexual reproduction, the reproduced genome is never an exact reproduction of one of the parents' genome, but a combination of chromosomes from both. Different offsprings of the same parents are also different from one another, because they do not get the same chromosomes from each parent. More variance is introduced through the process of 'crossing over': during meiosis (the process whereby the cell nucleus divides to create nuclei containing half the chromosome number of the parental nucleus), there is an interchange of sections between pairing homologous chromosomes. As a result, *no* chromosome of an offspring is identical with any of his parents' chromosomes. Finally, even variants that are a result of a chance mutation, are in a sense "designed" to appear: the gene copying processes contain repair mechanisms, that correct mistakes in DNA base pairings by replacing inappropriate or chemically altered nucleotides with the normal ones. Yet, not all mutations are corrected, as the repair mechanism itself makes mistakes. Why do they make such mistakes? Luria, Gould and Singer (1981) answer that 'Perhaps a "leaky" system that repairs most damaged DNA but does not fix *all* mutations as they occur has proved advantageous in evolution.' (*ibid*, p. 274). If a repair system has been selected for making mistakes, it is doubtful whether these can be treated as mistakes[4].

What all this points out is that variants are not just failures of a copying device to do a good job. Instead, there are mechanisms that are selected for producing variants; variants are *supposed* to appear, just as copies are. It is not difficult to see what selectional advantage such mechanisms of variance have: or-

[4]See (Luria, Gould and Singer 1981) for a description of the various mechanisms that introduce variance in the sexual reproduction process.

ganisms are living in a constantly changing environment, they are in competition with other organisms for different resources, and are threatened by predators. If they are to improve their adaptation to existing conditions, or adapt to changing conditions, they must change. Mechanisms are needed, therefore, that will ensure that enough variance is present in the replication processes.

7.2.4 The proper function of variants

Let us consider now what exactly is the proper function of those mechanisms of variance. Again, this proper function is relational—they are to produce genes that are *different* from the parental genes. In what way the reproduced genes will be different, and accordingly, what new effects they may have, is random. Since we are assuming a Darwinian (rather than Lamarkian) view of evolution, the variation that is introduced is not a directed one, i.e. the variants that will occur will not all, or even most, be better adapted than their parents. Nevertheless, because of the fact that *some* of the changes will have a survival value for the organisms in which they occur, mechanisms that introduce variance are selected for. That is, organisms that contain them will proliferate better than organisms with no variance in the reproduction process.

What is the derived proper function of the variants themselves, as products of the variance mechanisms? To answer this, we need to identify the proper function of the variance mechanisms 'that lies *beyond* the production of these adapted devices themselves'. That proper function is to introduce adaptive changes to the devices which they reproduce. Accordingly, the derived proper function of any particular variant is to (somehow) make the device it is a reproduction of better adapted. I have to admit that I am not entirely happy with this formulation—it seems circular to say that a device has the proper function of being adaptive in some general way. I cannot find, however, a better

formulation and I do believe that the underlying idea must be true: the reason why mechanisms of change proliferate must be because there is an advantage in having them. And this advantage must be the adaptive changes they introduce to the devices that they reproduce.

Perhaps the formulation can be improved, but one thing is clear: variant devices that proliferate themselves due to their new properties, i.e. those variant devices that will later acquire a new proper function, are the ones that performed their proper function *as variants*. It is those variant devices which proliferate that make the variance mechanisms themselves proliferate. This gives us the basis for discriminating a unique subclass among all variants that appear on the evolutionary scene, that is, the subclass of variants that have acquired a new proper function. We may now explore whether anything systematic can be said of this subclass of variants, in terms of their relations to the reproductively established families from which they originated.

Luria, Gould and Singer (1981) point out the following requirement that is assumed by the Darwinian theory of evolution: 'Darwin insisted that evolution must generally occur in small steps, through the successive incorporation of small changes by natural selection; adaptation is built gradually by the accumulating power of natural selection, not suddenly by large and lucky accidents' (*ibid*, p. 583). In Millikan's terms, this means that in the Normal case, i.e. in most cases where variant forms *did* become established, their new form and accordingly their new effects, were closely related to that of their ancestral devices. This is hardly surprising as most, if not all functioning devices are interdependent for their functioning on many other functionally related devices. Unless there is such an unlikely large and lucky accident where all those interdependent devices change at once, in a way that keeps them functionally synchronized, no large change to any one device can be an adaptive change. Conversely, any adaptive changes that occur must, Normally, be relatively small. The new form and

proper function that a device may acquire, therefore, while random in the above sense of not being directed, must be constrained by the form and function of the device from which it has originated.

If we consider again the question of what is the *direct* proper function of variants *as new devices*, we still must say that they do not have one. The new effects they might have due to their varied form are not their direct (or derived) proper function. Moreover, *what* these effects are going to be, is not determined by their history; the variants (as new devices) are not designed to do anything in particular. They were only designed, as variants, to be different. But while what their effects will be is not *determined* by their history, it is constrained by it in the case of variants that will proliferate. As explained, variants will have to be quite similar to their ancestral devices if they are to proliferate, and so will their effects have to be closely related to the functions of those devices. And when these effects are repeated enough times to make them a newly acquired proper function of this new reproductively established family of devices, these new proper functions will be closely related to the proper function of their original ancestral devices.

From the above description of the different mechanisms of change, we can see that nature has 'hit upon the idea' that introducing systematically constrained variation is advantageous over random variation. Instead of depending on pure chance mutations as the source of variance, mechanisms have developed that maintain variance while ensuring that most variants produced are similar enough to the parent genes to have a good chance of being functional. This, of course, doesn't mean that most variants will be *better* adapted than their parent genes. Quite the contrary: since the particular changes are only constrained, but not directed, only a minute proportion of the variants will in fact proliferate and acquire new proper functions. But this proportion will be greater than the proportion of chance mutations that will proliferate, giving those mecha-

nisms that maintain a constrained level of variance a selectional advantage over mechanisms of pure chance variance.

To summarize: we have explored the significance of the evolutionary history of variant devices in determining both their derived and direct proper functions. A distinction has been made between a variant device's being a variant as such, and its being a new device. In the case of devices that have specialized processes that control their reproduction, we have seen that the products of such processes do have an evolutionary history, in the sense that they belong to a certain first order reproductively established family (whether they share with it the relevant reproductively established properties or not). This determines their direct proper function: it is the same as the direct proper function of all other members of that reproductively established family, whether they can perform it or not.

As products of a specialized reproduction process, with relational proper functions, reproduced devices also have derived proper functions. Looking at genetic reproduction processes, we have seen that some mechanisms that control the reproduction process have the relational proper function of producing genetic material that is identical to the parent genetic material in some respects. As products of *those* mechanisms, genes have a derived proper function that is identical to their direct proper function. Other mechanisms that control the reproduction process have the relational proper function of producing genetic material that is varied (in a non-specific way) relative to the parental genetic material. Genes have, as products of those mechanisms, the derived proper function of introducing adaptive changes to the form, and thus functioning, of the parent genes. Variant genes that will perform their derived proper function as variants will acquire new direct proper functions that are closely related to the direct proper function of the parent genes. Exactly what direct proper function a variant will acquire is under-determined by its evolutionary history,

which determines its derived proper function. Limited generalizations can be made, however, about what direct proper function a variant can acquire, *if it does acquire one.* These generalizations can be stated in terms of the new proper function's relation to the gene's present direct proper function as a member of an existent reproductively established family. Such limited generalizations will apply to the restricted subclass (however small) of variants that do perform their proper function as such.

7.3 Metaphor as a semantic mechanism of variance

Millikan's explanation of what it is for a linguistic device to have its own meaning, recall, is the following: it is its having a direct, stabilizing proper function, Normal mapping conditions for the performance of this function (in the case of referential or denotative language devices) and public intensions, as far as there are such. These aspects of meaning are characterized independently of the intentions of any particular user of a linguistic device. The speaker's intentions determine only the derived proper function of a linguistic device token, and do not directly effect its *own* meaning, as a public device. Does a fresh metaphor have a public 'metaphorical meaning'? Is there something it is supposed to do, in some sense, that is different from its direct, stabilizing function, and yet independent of its utterer's intentions? I will try to answer this question in the light of the analogy between metaphors and biological variants.

7.3.1 The proper function of fresh metaphors

Since the meaning of a linguistic device is determined by its history, according to Millikan, the first question we need to address is what kind of evolutionary

history does a fresh metaphor have. Is a fresh metaphor a new linguistic device with a new meaning, or is it just another use of an existing linguistic device? The tension between these two aspects is brought out in Davidson's words:

> 'If we are to think of words in metaphors as directly going about their business of applying to what they properly do apply to, there is no difference between metaphor and the introduction of a new term into our vocabulary... What has been left out is any appeal to the original meaning of the words. Whether or not metaphor depends on new or extended meanings, it certainly depends in some way on the original meaning...'.
>
> (Davidson 1984, pp. 248–9)

Davidson, we have seen, thinks that the original meanings are enough, and there is no need to consider metaphor as meaning anything new or extended. Both Davidson and Millikan, pointing at the different sides of the same coin, impose on us an all or none decision: either a mutant or a metaphor are new, like a new term in the vocabulary or a cosmic accident, or they aren't. What I have tried to do in the former section is to reject this all or none approach, and show that in the case of biological variants, their history as one device has significance in determining their future as another. This, I believe, is the sensible way of treating metaphor.

As was done with biological variants, we should distinguish between a metaphor *as such*, and a metaphor as a particular linguistic device with specific new effects. Saying of an expression that it is a metaphor implies two things: it is a metaphor only if it is not just a new term but bears clear relations to an existing set of reproductively established families of language devices (as Davidson, in his own terminology, insists); but to be a metaphor it must also have effects that are new and do *not* accord with its proper function as a reproduced device. An expression is a metaphor, as such, if it is a replication, through imitation, of some existing language devices. Also, it must be an expression that has some

effects on the hearer that are different from the effects that it is supposed to have as a reproduction of certain language devices (though being an imitation with variant effects is not sufficient for being a metaphor). I will postpone the discussion of metaphors as specific new devices until later.

A short diversion: notice that there is a difference between the case of being a biological variant, and that of being a linguistic variant. The first implies a change in the form of a device, that brings about a change in the effects that a device has. The second, instead, is defined directly by its having different effects, and does not assume a change in form; true enough, words that acquire new meanings do not change their form. This difference derives from the fact that for many biological devices, their form is directly relevant to their functioning, whereas in language devices the connection between a word and its function seems arbitrary. This may cause some confusion, as to whether language devices that acquire new proper functions are new devices, or just old devices with new functions. But there is no real problem here: a device that acquires a new function is a new device, even if its form hasn't changed. Devices are individuated by their proper function, not their form. A change in a device's functioning may occur because of variance in its form, but it may also take place because of variance in the conditions in which it is operating.

It should also be noted that there is a sense in which the form of language devices, and intentional icons in general, *is* relevant to their functioning, given that their functioning depends on particular response mechanisms. One could think, for example, that the fact that the pigment arranging device in the chameleon produces skin colors that are identical to its surroundings is necessary to its functioning as a camouflage device, whereas the fact that 'dog' initiates in hearers the identification of a dog is arbitrary. But the functioning of the pigment arranging device in the chameleon as a camouflage device depends on the structure of its predetor's sight mechanisms, e.g. it would be inefficient if the chameleon's

predators were blind, but used heat detectors. Similarly, a beaver's "arbitrary" way of signaling danger by a tail-splash is dependent on other beavers' response mechanisms that interpret it as a warning signal; and the "arbitrary" use of 'dog' to initiate in hearers an identification of a dog depends on their acquired response to 'dog' in this way.

Back to the discussion of metaphors, as such. A metaphor is defined relative to the reproductively established family of which it is a reproduction. Does it have a proper function as a metaphor? Given that it is a product of specialized imitation processes, I would like to claim that it does. Those imitation processes are 'rewarded' both for producing devices that perform their direct proper function, and for producing devices that have other adaptive effects. Thus, an expression has, as a product of these mechanisms, two different derived proper functions. One is to perform that function which is the direct proper function of the reproductively established family of devices to which it belongs. The other is to introduce some other effects that will be adaptive. In other words, these specialized processes produce expressions with variant effects not just as a malfunction, but as a proper function of those devices. Communication devices that were too stable would be just as inefficient as any other device that never varied. Dawkin's main idea in his book 'The Extended Phenotype' (Dawkins 1983) helps to see this point. He argues that we should treat the effects of different communicative systems on other organisms as phenotypic effects of some genes, in the same way as we treat the construction of specific devices (e.g. hearts, eyes, etc.) in the same organism as the phenotypic effects of other genes. We have seen the importance of variance to the proliferation of the genes that control the latter type of phenotypic effects, and presumably variance is just as essential to the proliferation of communication devices.

A difference in flexibility is perhaps one of the important things that distinguishes animal sign systems from human natural languages. The first are very

stable, because of being "wired in" to the organisms that use them and can only change through long evolutionary processes. The expressibility of those sign systems is therefore very limited. Human languages are much more flexible, because they depend on plastic learning mechanisms rather than being "wired in". The availability of mechanisms that underlie semantic change such as metaphor, it seems, may account for the human languages' much greater expressive powers (that, over and above the generative nature of human languages that is usually mentioned in this context).

That language devices do vary in their effects is an obvious fact, and metaphor is a clear case of such variance. What the analogy with the biological variants brings out is that metaphors, as such, should be treated as a functional device rather than a malfunctioning one, as Millikan would have it. Their function is derived from the function of the mechanisms in speakers that produce them, and it is to introduce adaptive changes in the devices of which they are varied reproductions (varied, this time, in their effects rather than their form).

7.3.2 Separating functioning from non-functioning metaphors

The most crucial implication of the analogy made between biological variants and metaphors, is the following. If we are to make any valid generalizations about how metaphors function (that is, how they introduce adaptive changes), we must restrict those claims to the subclass of metaphors that *do* function. In the case of indicative sentences, we have seen that Millikan's claims about their Normal mapping rules need only apply to those which perform their proper function of creating true beliefs in the hearers. Similarly, any claims we will want to make about the Normal character of metaphors should only apply to those metaphors that perform their proper function as such. Now, it was pointed out that biological variants which did perform their proper function as variants

were the ones that had proliferated. The analogy suggests that metaphorical expressions that had become established, and had acquired new meanings, were the ones that performed their proper function as metaphors. Generalizations that are to be made about metaphorical effects need only apply therefore to this subclass of metaphors. The important thing about this line of reasoning is that it provides a principled reason for restricting generalizations to this subclass, rather than making claims that are, just as a matter of fact, true of an arbitrarily limited range of metaphors.

Let us go back and consider one of Cooper's criticisms against Searle that was discussed in Chapter 1. He says that '...very few of the examples [which Searle uses] are, by the lights of the standard theory[5] itself, examples of metaphor at all. Almost to a man, they are ones of 'dead' metaphor' (Cooper 1986, pp. 68–9). He then adds, '...it is an illegitimate and one-sided diet of examples which lends plausibility to the standard view' (*ibid*, p. 70). This criticism, which is valid against Searle's standard view, is answered here by turning the problem on its head. I shall restate, and further develop, the answer to this criticism which I have suggested in Chapter 1.

Searle's principles of interpretation apply perhaps only to 'dead' metaphors, and this is reflected in his choice of examples. Moreover, it *is* doubtful whether any interpretation principles can be found that will apply to *all* metaphors. But in the account presented here, principled relations between metaphorical expressions' original meanings and their new meanings acquired through metaphorical use only *need* to apply to 'dead', or established metaphors (from this perspective, it sounds odd to call established metaphors 'dead', as they are exactly the ones that survived, rather than died). Searle could be justified, within the framework of this account, in saying that '...we need not feel apologetic about

[5]Cooper labels theories that identify metaphorical meaning with the speaker's intended meaning as 'standard theories' of metaphor. Searle's is a prototypical example of the standard theories.

the fact that some of our examples are trite or dead metaphors. Dead metaphors are especially interesting for our study, because, to speak oxymoronically, dead metaphors have lived on. They have become dead through continual use, but their continual use is a clue that they satisfy some semantic need' (Searle 1979, p. 83). But, the examples of trite or dead metaphors are only that—examples of trite or dead metaphors. They do not exemplify *all* metaphors, just as, to use Millikan's example up-side-down, healthy hearts cannot exemplify diseased ones. Given, however, that his account is meant to apply to all metaphors, Searle *should* feel apologetic about his examples.

Searle's principles of interpretation need to be re-examined in the light of the account of metaphor suggested here. Perhaps they can serve as a good starting point for making valid generalizations about how metaphors perform their proper function as such, the function of introducing semantic change. These principles should not be treated, however, as 'principles of interpretation' in the sense of being claims about how people go about interpreting metaphors, and even less as principles for finding out *intended* meanings. What these principles might show, is what relations do in fact hold, in the Normal case, between original meanings and new ones acquired through metaphorical use (more on this shortly).

We may find out, however, that Searle's principles are still too restrictive, and do not apply to most cases of metaphorical change, even when those changes become established. Maybe the generalizations that can be made about systematicity in metaphorical change are quite limited. Hesse, acknowledging these difficulties, still thinks that such generalizations would be of much interest: 'What *would* be of philosophical interest would be a successful account of how models and analogical frameworks work in scientific inference, and by extension, how metaphor works in discourse. It may well turn out that nothing but broad generalities are possible in such an account, supplemented by detailed study of

particular cases' (Hesse 1987, p. 305). As we've seen in Chapter 3, Hesse's understanding of the question 'how metaphors work?' comes close to asking how do metaphors effect changes in meaning.

The account suggested here is in full agreement with Hesse's view that the study of metaphor should be the study of systematicity in semantic change. It has, however, the following advantages over Hesse's account: first, it explicitly makes a distinction between metaphors that do initiate semantic change and those that do not, which Hesse fails to do. If her account, like Searle's, is assumed to be applicable to all metaphors, it can be rejected for its empirical inadequacy: not all metaphors do introduce semantic change. Our present account not only makes the necessary distinction (that could be added, ad-hoc, to Hesse's account), but also provides principled reasons for why an account of metaphor can legitimately restrict itself to those metaphors that do introduce semantic change. Second, Hesse's account doesn't show why semantic change should be a proper part of a semantic theory. By contrast, it follows naturally from the evolutionary approach that meaning should be considered diachronically. While it is not a-priori evident what semantics should deal with, it is clear that variance and change are a proper part of evolution theory, and thus should be a part of a semantics based on it.

7.3.3 Identifying the Normal characteristics of functional metaphors

I shall now try to sketch some generalizations that can be made about metaphors that become established. First, it must be stressed that there are many factors, specific to each particular metaphor, that determine whether it is to become established or not. The following is therefore not intended as an attempt to specify sufficient conditions for a metaphor to become established, and I am certain that it also falls short of specifying the necessary conditions. I hope,

however, that it will give some indication of the kind of generalizations that can be made about semantic changes through metaphorical use.

One Normal character of metaphors that become established, and the one most relevant to their semantics, is that their new, acquired meanings (i.e. their Normal mapping value) should retain certain systematic relations to their original meanings. Those are the kind of relations that Searle has tried to capture by his interpretation principles, and this is where I think his work can be of value. For instance: the acquired meaning of a referential term R could be such that it is, by definition, a characteristic of P, its original meaning. Searle's example: 'giant' can acquire the meaning *very big*, because giants are, by definition, very big. Or, the acquired meaning of a referential term could be some salient or well known property of its original meaning. Searle's example: 'pig' can acquire the meaning *filthy, gluttonous and sloppy.* Similarly with Searle's other principles and examples.

At the risk of being repetitive, let me stress again how the role of Searle's principles in this account is different from their role in his own account. I am not saying that *all* metaphors do maintain these relations. I am not arguing that all metaphors that are understood by hearers maintain these relations. I am not claiming that when metaphors are understood, it is because speakers and hearers 'calculate' over their literal meanings using these principles. All I am saying is that at best, these relations are maintained in most cases between an expression's new meaning that was acquired through repetitive metaphorical use, and its original meaning. Also, these relations are taken as constraints on possible acquired meanings, not as strict determinants of these meanings.

Why should such relations be expected to hold between an expression's new meaning and its original meaning? In the case of biological devices, we expected some systematicity between a device's original and acquired proper functions

because of the interdependency of a device's functioning on other functionally related devices. Here the reason for expecting such (limited) systematicity has to do with the common psychological processes that underlie interpretation. Metaphors that will have effects that meet such constraints will stand a better chance of becoming established because they will be more uniformly interpreted, than metaphors that depend for their interpretation on a specific context. They can be repeated, therefore, in a wider variety of occasions, and will tend to acquire, through repetition, a new meaning.

Which leads us to the next Normal character of metaphors that perform their proper function of introducing semantic change. When a fresh metaphor is used, its effects on the hearer will in most cases be the effect that was intended by the speaker, in the case of metaphors that become established. In Millikan's terms: a metaphor's actual effects on the hearer will be the same as the derived proper function of that expression, that is determined by the speaker's intentions. In other words, metaphors that become established will be, in most cases, understood as intended even before they have acquired a new meaning, that is, even on the first occasion of their being used. This follows from the consideration that for a linguistic device to acquire a direct proper function, it must have stabilizing effects that will encourage speakers to keep using, and hearers to uniformly respond to, these expressions. If a speaker who uses a metaphor is not understood, neither he or his hearer will tend to continue using that metaphor. Again it should be stressed that this doesn't mean that *most* metaphors will in fact be understood as intended. It only means that in most cases *where a metaphor becomes established*, it will be understood as intended even before it has become established.

Toulmin, who also compares linguistic change to evolutionary processes, makes an analogy that helps support the above point. His claim, made about language in general rather than about specific language devices, is the following:

'...members of the various speech-communities employing it [i.e. a language] should be capable of making themselves sufficiently intelligible to one another. (This corresponds to the requirement of natural interfertility, which plays a corresponding part in the popular definition of an organic 'species'.)'.

(Toulmin 1972, p. 343)

This point applies not only to languages or species in general, but to particular devices as well. A biological variant device can only proliferate if it doesn't interfere with the interfertility of an organism in which it appears with other organisms of its kind. Similarly, a variant linguistic device, a metaphor, can only become established if it remains intelligible, even though it deviates from the standard.

A further hypothetical characteristic of metaphors is that the semantic changes that they will bring about will be such that they will enable people to communicate (i.e. convey information, or cause specific actions) about things that were not communicable given the existing linguistic apparatus. In other words, changes will tend to be adaptive when they close 'semantic gaps'. As Searle says, 'It is often the case that we use metaphor precisely because there is no literal expression that expresses exactly what we mean' (Searle 1979, p. 114). It is this property of metaphors which makes the languages of humans, that cope with metaphorical variance, more flexible and with larger expressive powers than animal sign systems.

7.3.4 Metaphorical meaning

Up until now I have discussed the proper function of metaphors as such, and claimed that their derived proper function singles out metaphors that have become established as the appropriate domain of investigation about metaphors that have performed their function as such. Let me turn now to the other aspect

of fresh metaphors—their first occurrence as a specific new device, with new effects. Does a metaphor have a direct proper function, or meaning, as a new device? It doesn't, and it couldn't, by definition. There is nothing in particular that a fresh metaphor is supposed to do as a new device. But then, what does the notion of the 'metaphorical meaning' of an expression apply to?

As I've pointed out in the discussion of biological variants, which effects a variant should have is not determined by its history. What is *constrained* by its history, however, though still under-determined, is which effects have the chance of proliferating; and being a variant as such, it *should* proliferate. Similarly with metaphor: there is no particular way a fresh metaphor should be interpreted. On its first appearance, as Cooper has put it, it belongs to the world, not to its utterer, and no particular interpretation can be wrong in any sense. For an interpretation to be wrong, there must be norms against which it is compared; in this account, norms are set by the history, and a fresh metaphor (as a new device) has no history. Moreover, this account agrees, to a certain extent, with Davidson's claim that metaphors are used many times for their non-cognitive effects, i.e. those effects that do not involve the conveying of any propositional content.

Davidson's account, however, has no power to illuminate the fact that (some) metaphors *do* initiate semantic change. As I pointed out in Chapter 2, it is a major weakness of Davidson's account that it doesn't explain how non-cognitive, non-semantic effects mature into semantic changes. The account presented here overcomes this problem. A fresh metaphor may have non-cognitive effects, as well as cognitive ones; on its first appearance, none of these can be distinguished as a metaphor's proper function. But its history as a semantic device, in particular its proper function as a semantic variant, constrain (though still under-determine) which of its effects may proliferate and turn out to be its direct, stabilizing proper function after enough reproductions.

What is, then, a fresh metaphor's 'metaphorical meaning'? I suggest that this notion should apply to the constrained, but perhaps open-ended, range of semantic effects that a metaphor might have, that stand a chance of becoming established. As I have pointed out, studying the relations between expressions with meanings that were established through metaphorical use and the original meaning of those expressions may serve to discover the nature of those constraints that define an expression's metaphorical meaning, in the above sense. The choice to concentrate on the semantic effects of fresh metaphors, I have argued, is not an arbitrary one but derives from analyzing the function of metaphors through the analogy with biological variants.

To substantiate this claim, I must explain what distinguishes a semantic effect of an expression from other effects. Millikan explicates the notion of semantic content in terms of the *sense* of an expression, i.e. its Normal mapping rules. To recall, she says that the sense of a referential or denotative term is that aspect of the world to which it corresponded in most cases where that term has performed its proper function. We have a problem here, because we are trying to define a semantic content for a new device, with no history of Normal functioning. We may overcome this problem, however, by considering how a linguistic device's semantic content becomes established. The proper function of indicative and imperative sentences is to produce, respectively, a true belief in the hearer, or an action. These sentences map onto the world according to certain rules when they perform those functions Normally. We may distinguish the semantic effects of a fresh metaphorical expression, then, not by its Normal conditions for proper functioning (because there is no Normal explanation for those effects), but by characterizing the semantic effects as those that produce a belief or intention to act in the hearer. It is this kind of effects that, when becoming established as an expression's proper function, will have Normal conditions for their functioning that will involve the relevant kind of mapping rules with the world.

One question that remained open so far is what makes an expression into a metaphor. An expression may be considered a metaphor for two independent reasons: either for its being intended metaphorically, or for being taken metaphorically. An expression is intended metaphorically if its utterer purposes that it should have effects on its hearer that are other than its direct, stabilizing proper function. An expression is taken metaphorically if it in fact has effects that are other than its direct, stabilizing proper function.

This is clearly not enough: the above definition will include cases of lying, or misunderstanding, or irony, etc. To distinguish metaphor from all other non-standard intended uses or effects, we must say in what way a metaphor's effects are different from all those other non-standard effects. Doing this is problematic. It may seem that this account offers a straightforward answer, given the availability of the notion of metaphorical meaning as the range of effects that are within the limits set by the Normal characters of metaphors that have become established. We may say that an expression is a metaphor if its intended, or actual effects are within those limits. This answer is not satisfactory, however, for at least two reasons. First, we do in fact treat expressions as metaphors even if they never become established—just as we do treat as indicative sentences many that are not true. The whole idea of separating functional from non-functional metaphors is undermined if we exclude non-establishable metaphors from metaphorhood. Second, I have said that many metaphors only have non-cognitive effects, and those are still metaphors even if they cannot underlie semantic changes (this, in fact, is a special case of what I have just considered in my first objection to the suggested answer). I do not have, however, a better solution to this problem of distinguishing metaphorical effects from others, and I will have to leave it open.

7.3.5 Metaphorical clusters

In Chapter 1 I have mentioned Lakoff and Johnson's claim that many metaphorical expressions tend to become established in semantically related clusters (see p. 30). I would like to point out how it follows from the analogy between metaphors and biological variants that such clusters should be created. As I have explained in the section about biological variants, biological devices usually depend for their functioning on many functionally related devices—this is what delimits the scope of functional changes that can take place through variance. Another result that this functional dependency has, is that a change to one device will re-define the scope of changes that will be adaptive in all the other devices that are functionally related to it.

Cooper says that 'it is the exception, rather than the rule, for established metaphorical expressions to have become established singly' (Cooper 1986, p. 130). This remains true if you replace 'metaphorical expression' with 'biological variants'. Once one change takes place in a biological system of devices, many related changes will tend to follow. The classic example is that of the chain of changes that took place as our ancestors started walking on their hind legs, freeing their hands for other activities. This was followed by a change in the position of the thumb in relation to the other fingers, a change in the sensitivity of the manipulation of the hands, etc. Similarly, the change from amphibian life to terrestrial life was followed by changes in the limbs, the breathing systems, and so on. A similar situation occurs with metaphors. Once one change takes place, it changes the context for many other related words which enables their variant effects to become established. This could explain the phenomenon of metaphorical clusters: a change to one words alters the range of possible metaphorical meanings, i.e. the range of effects that are likely to become established, for many other semantically related words.

In a more general way, the approach presented here emphasizes the relevance of inquiries such as Lakoff and Johnson's into the structure of established metaphors, to the understanding of the phenomenon of metaphor in general. Such an inquiry may seem redundant to accounts of metaphor that exclude established metaphors from their programme. In this account, instead, their study may be given an essential role. In particular, the structure they have revealed between established metaphors suggests that an underlying systematic structure can also be revealed between established metaphors and their common source, i.e. their original, "pre-metaphoric" meaning.

7.4 Conclusion

In the first part of the thesis we have encountered three accounts of metaphorical meaning, each pointing out important observations, and providing valuable insights on the subject. Each of the writers was supporting his or her claim by implicitly assuming, or explicitly arguing for, one of the contradicting possibilities—either that metaphor was systematic, or that it wasn't. Though contradictory, all claims had their intuitive appeal. My hope is that the view presented in this last chapter, managed to capture the significant insights made by the three philosophers, yet make their views compatible, and resolve the tension between the apparently opposed intuitions about systematicity in metaphor on which they were relying. Let me specify, by way of summary, how this was done.

Searle has assumed that a systematic relation holds between the metaphorical meaning of expressions and their literal meaning. Cooper pointed out, though, that the examples Searle had used in analyzing this systematicity were all of 'dead', or established metaphors. I have gone along both with Searle's assumption and with Cooper's criticism of it. I have argued that the system-

aticities only apply in cases of established metaphors, but that Millikan's theory together with my extension of it show why such limited application was to be expected, and accepted. The approach presented here agrees with Searle's claim that established metaphors are of special interest, and endorses his view that metaphors have the role of closing 'semantic gaps'. I rejected, though, Searle's implicit thought that his theory could apply to metaphors in general, rather than to the sub-class of metaphors that become established. I also denied his attribution of metaphorical meaning to the intentions of the speaker; more generally, I followed Millikan in preferring an approach that maintained a certain gap between the notion of meaning, and the intentions and comprehension processes that accompany particular linguistic exchanges. For that reason, I also rejected Searle's claim that what he was offering was a rational reconstruction of comprehension processes, and argued that this sort of account was redundant.

Davidson has put forward several arguments for denying the idea that metaphors have a special meaning. First, he claimed, metaphors do not have a specific cognitive content; second, their significance was in their non-cognitive effects, which are not amenable to a semantic treatment; finally, metaphors are creative achievements both in their making and in their understanding, and cannot therefore be captured by systematic rules for interpretation. I agree with Davidson's latter claim, and with the general thrust of his view, that a fresh metaphor doesn't have a meaning. This claim was captured in the present account by the claim that a metaphor, as a new linguistic device, cannot be attributed with a proper function, or meaning. It was further accepted that even in the case of metaphors that become established, their original meaning doesn't *uniquely* determine their metaphorical meaning. I argued against his other views, however: some metaphors do acquire new meanings, so they must have a (non-specific) cognitive content to begin with; and metaphors surely have significant non-cognitive effects, but the fact that this has nothing to do with semantics doesn't show that a semantic treatment of metaphor is not possible or

appropriate. More generally, I think that Davidson's approach which highlighted non-systematicity in metaphor was correct, given the synchronic approach to meaning that he has adopted. Millikan's view, however, of meaning as a matter of the history of language devices, offers another perspective on metaphor, from which the systematicities that lack in Davidson's account can be extracted.

Finally, there was Hesse's approach to metaphor. The view presented here pretty much agrees with her opinion, that an account of metaphors should concentrate on their role in initiating semantic change, however limited are the generalizations that can be made regarding this issue. This account, however, makes several advances over her argumentation. It does so, first, by relying on Millikan's more detailed and precise notion of meaning; second, by the explicit discrimination it makes between those metaphors that do underlie semantic change, and those which do not; and finally, by its reliance on a theory that provides principled grounds for considering semantics diachronically.

In trying to reconcile the three opposing views, I have followed the wisdom offered by this Jewish fable: when two members of the community had a quarrel, they came to consult their rabbi. The first entered the rabbi's room and presented his claims, to which the rabbi responded by nodding, and saying 'you are certainly right'. The first left happily, and the second came in. He pleaded his case, to which the rabbi again responded: 'you are certainly right'. When the second had left, the rabbi's wife entered and perplexedly asked: 'excuse me for overhearing your conversation, but how could you have approved of both these people, when they were arguing for entirely opposing view?' The rabbi thought for a moment, and then answered: 'I must say that you are certainly right as well'.

References

Bennett J.F. (1976). *Linguistic Behaviour*, Cambridge: Cambridge University Press.

Black M. (1955). 'Metaphor', *Proceedings of The Aristotelian Society*, vol. 55: 273–294.

Black M. (1978). 'How Metaphors Work: A Reply to Donald Davidson', in Sacks S. (ed.), *On Metaphor*, Chicago: The University of Chicago Press.

Cooper D. (1986). *Metaphor*, Aristotelian Society Series, Volume 5, Oxford: Basil Blackwell.

Davidson D. (1984). 'What Metaphors Mean', in *Inquiries into Truth and Interpretation*, Oxford: Clarendon Press.

Davidson D. (1986). 'A Nice Derangement of Epitaphs', in LePore E. (ed.), *Truth and Interpretation: Perspectives on The Philosophy of Donald Davidson*, Oxford: Basil Blackwell.

Dawkins R. (1983). *The Extended Phenotype*, Oxford: Oxford University Press.

Dawkins R. (1989). *The Selfish Gene*, Oxford: Oxford University Press.

Dummett M. (1986). ''A Nice Derangement of Epitaphs': some comments on Davidson and Hacking', in LePore E. (ed.), *Truth and Interpretation: Perspectives on The Philosophy of Donald Davidson*, Oxford: Basil Blackwell.

Fodor J. (1987). *Psychosemantics*, Cambridge, Mass.: Bradford Books/MIT Press.

Fodor J. (forthcoming). 'A Theory of Content', in Rey G. and Loewer B. (eds.), *Meaning and Mind: Fodor and His Critics*, Blackwell.

Grice H.P. (1957). 'Meaning', *Philosophical Review*, vol. 66: 377–388.

Grice H. P. (1975). 'Logic and Conversation', in Cole P. and Morgan J. (eds.), *Syntax and Semantics 3: Speech Acts*, New York: Academic Press.

Goodman N. (1978). 'Metaphor as Moonlighting', in Sacks S. (ed.), *On Metaphor*, Chicago: The University of Chicago Press.

Harris J. (1976). 'Comprehension of Metaphor: A Test of the Two-Stage Processing Model', *Bulletin of the Psychonomic Society*, vol. 8: 321–324.

Hesse M. (1966). 'The Explanatory Role of Metaphor', in *Models and Analogies in Science*, Notre Dame, Indiana: University of Notre Dame Press.

Hesse M. (1987). 'Tropical Talk: The Myth of The Literal', *Proceedings of The Aristotelian Society Supplementary Volume LXI*: 297–311.

Lakoff G. and Johnson M. (1980). *Metaphors We Live By*, Chicago: University of Chicago Press.

Luria S. E., Gould J. S. and Singer S. (1981). *A View of Life*, Menlo Park, California: The Benjamin/Cummings Publishing Company, Inc.

Millikan R. G. (1984). *Language, Thought and Other Biological Categories*, Cambridge, Mass: Bradford Books/The MIT Press.

Millikan R.G. (1989). 'Biosemantics', *The Journal of Philosophy*, vol. 86: 281–297.

Ortony A., Schallert D., Reynolds R. and Antos S. (1978). 'Interpreting Metaphors and Idioms: Some Effects of Context on Comprehension', *Journal of Verbal Learning and Verbal Behaviour*, vol. 17: 465–478.

Riquere P. (1978). 'The Metaphorical Process as Cognition, Imagination, and Feeling', in Sacks S. (ed.), *On Metaphor*, Chicago: The University of Chicago Press.

Rorty R. (1987). 'Unfamiliar Noises: Hesse and Davidson on Metaphor', *Proceedings of The Aristotelian Society Supplementary Volume LXI*: 283–296.

Searle J. (1979). *Expression and Meaning*, Cambridge: Cambridge University Press.

Sperber D. and Wilson D. (1986). *Relevance*, Oxford: Basil Blackwell.

Toulmin S. (1972). *Human Understanding, Volume 1: The Collective Use and Evolution of Concepts*, Princeton, New Jersey: Princeton University Press.